



STIC Search Report

Biotech-Chem Library

STIC Database Tracking Number: 108396

TO: Sheridan Swope
Location: CM1/12S12&10D)1-
Art Unit: 1652 12D12 / 10D01
Tuesday, November 25, 2003

Case Serial Number: 09/607745

From: Barb O'Bryen
Location: Biotech-Chem Library
CM1-6A05
Phone: 308-4291

barbara.obryen@uspto.gov

Search Notes

STIC-Biotech/ChemLib

10 8396

From: Swope, Sheridan
Sent: Monday, November 17, 2003 12:48 PM
To: STIC-Biotech/ChemLib
Subject: 09607745

For 09607745, pls do the following searches and alignments:

Search:

SID 2 against the NT and AA data bases.

SID 9 against the NT and AA data bases.

Align SID 2 and 9 with:

EMBL Acc# AF179224

EMBL Acc# AF216312.

EMBL Acc# Q9NRS4

GENSEQ Acc# AAY99417

CRFF

THANKS!!!

Sheridan Swope, Ph.D.
Patent Examiner, AU 1652
Recombinant Enzymes
sheridan.swope@uspto.gov
703-305-1696 (voice)
703-308-3014 (FAX)
Mailbox: CM1 Rm10D01
Office: CM1 Rm12D12

Searcher: _____
Phone: _____
Location: _____
Date Picked Up: _____
Date Completed: _____
Searcher Prep/Review: _____
Clerical: _____
Online time: _____

TYPE OF SEARCH:
NA Sequences: _____
AA Sequences: _____
Structures: _____
Bibliographic: _____
Litigation: _____
Full text: _____
Patent Family: _____
Other: _____

VENDOR/COST (where applic.)
STN: _____
DIALOG: _____
Questel/Orbit: _____
DRLink: _____
Lexis/Nexis: _____
Sequence Sys.: _____
WWW/Internet: _____
Other (specify): _____

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 25, 2003, 13:51:12 ; Search time 0.001 Seconds
(without alignments)
378.015 Million cell updates/sec

Title: US-09-607-745-2
Perfect score: 2342
Sequence: 1 MDPDSQPLNSLDVKEPLRKP.....VYTKVSAYLNWYNWKAEL 435

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2 seqs, 869 residues

Total number of hits satisfying chosen parameters: 2

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : aa.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2337	99.8	437	1 q9nrs4	TOIG of: q9nrs4
2	2297.5	98.1	432	1 aay99417	TOIG of: aay9941

ALIGNMENTS

RESULT 1
q9nrs4
TOIG of: q9nrs4 check: 316 from: 1 to: 437

TS4 HUMAN STANDARD; PRT; 437 AA.
Q9NRS4; Q9NRS4; 437 1 q9nrs4
16-OCT-2001 (Rel. 40, Created)
15-SEP-2003 (Rel. 42, Last annotation update)
Transmembrane protease, serine 4 (EC 3.4.21.-) (Membrane-type serine protease 2) (MT-SP2).
TPRSS4 OR TPRS33.
Homo sapiens (Human).
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NCBI_TaxID=9606;
[1]

SEQUENCE FROM N.A.
RC TISSUE=Pancratic carcinoma;
RX MEDLINE=20283276; PubMed=10825129;
RA Wallrapp C., Haehnel S., Mueller-Pillasch F., Burghardt B., Iwamura T., Ruthenburger M., Lerch M.M., Adler G., Gress T.M.;
RT "A novel transmembrane serine protease (TPRSS3) overexpressed in pancreatic cancer."
RL Cancer Res. 60:2602-2606 (2000).
RN [2]

SEQUENCE FROM N.A.
RA Smeekens S.S., Lorimer D.D., Wang E., Hou J., Linnevers C.;
RT "MT-SP2, a novel type II membrane serine protease expressed in trachea, colon, and small intestine: identification, cloning, and chromosomal localization."
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
RN [3]

SEQUENCE FROM N.A.
RP TISSUE=Pancratic;
RC MEDLINE=22388257; PubMed=12477932;
RX Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Klausner R.D., Collins F.S., Wagner L., Schenken C.M., Schuler G.D., Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K., Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F., Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L., Stapleton M., Soares M.B., Bonaldo A.F., Casavant T.L., Scheetz T.E., Brownstein M.J., Udell T.B., Toshiyuki S., Carninci P., Prange C., Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J., Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H., Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W., Villalón D.K., Muzny D.N., Sodergren E.J., Lu X., Gibbs R.A., Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A., Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G., Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C., Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallos D.E., Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).

FUNCTION: MAY PLAY A ROLE IN TUMOR PROGRESSION, METASTASIS FORMATION AND TUMOR INVASION.
SUBCELLULAR LOCATION: Type II membrane protein (Potential).
TISSUE SPECIFICITY: HIGH LEVELS IN PANCREATIC, GASTRIC, COLORECTAL AND AMPULLARY CANCER. VERY WEAK EXPRESSION IN NORMAL GASTROINTESTINAL AND UROGENITAL TRACT.
SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
SIMILARITY: Contains 1 SRCR domain.
SIMILARITY: Contains 1 LDL-receptor class A domain.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; AF179224; AAF74526.1; -
EMBL; AF216312; AAF31436.1; -
EMBL; BC011703; AAH11703.1; -
HSPF; P00763; IDPO.
MEROPS; S01.034; -
Gene; HGNC:11878; TMRSS4.
MIM; 606565; -
GO; GO:0016021; C:integral to membrane; NAS.
GO; GO:0006508; P:proteolysis and peptidolysis; NAS.
InterPro; IPR001314; Chymotrypsin.
InterPro; IPR002172; LDL receptor A.
InterPro; IPR001254; Ser. protease Try.
InterPro; IPR001190; Srcr receptor.
Pfam; PF000057; ldl_recept_a; 1.
Pfam; PF000089; tryptsin; 1.
PRINTS; PR00722; CHYMOTRYPSIN.
SMART; SM00192; LDLa; 1.
SMART; SM00202; SR; 1.
SMART; SM00020; Tryp_SPC; 1.
PROSITE; PS01209; LDLRA_1; FALSE NEG.
PROSITE; PS00068; LDLRA_2; FALSE NEG.
PROSITE; PS00420; SRCR_1; FALSE NEG.
PROSITE; PS00287; SRCR_2; 1.
PROSITE; PS0240; TRYPSIN_DOM; 1.
PROSITE; PS00134; TRYPSIN_HIS; 1.

Perfect score: 1600
Sequence: 1 MDSKSSQSRLLLLLVSN.....LNWIYNVWKAELSRHHHHH 292

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2 seqs, 869 residues

Total number of hits satisfying chosen parameters: 2

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : aa.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1266	79.1	432	1 aay99417	TOIG of: aay9941
2	1266	79.1	437	1 q9nrs4	TOIG of: q9nrs4

ALIGNMENTS

RESULT 1

aay99417
TOIG of: aay99417 check: 6343 from: 1 to: 432

ID AAY99417 standard; Protein: 432 AA.

XX XX

AC XX

DT XX

XX XX

DE XX

XX XX

KW KW

KW KW

XX XX

XX OS

XX XX

PN PD

XX XX

XX XX

PF PF

XX XX

XX XX

PR PR

PR PR

PR PR

PR PR

PR PR

PR PR

PR PR

PR PR

PR PR

PR PR

PR PR

PR PR

PR PR

PR PR

PR PR

PR PR


```

-!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
-!- SIMILARITY: Contains 1 SRCR domain.
-!- SIMILARITY: Contains 1 LDL-receptor class A domain.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; AF179224; AAF74526.1; -
EMBL; AF216312; AAF31436.1; -
EMBL; BC011703; AAH11703.1; -
HSP; P00763; IDPO.
MEROPS; S01.034; -
Genew; HGNC:11878; TMRSS4.
MIM; 606565; -
GO; GO:0016021; C: integral to membrane; NAS.
GO; GO:0006508; P: proteolysis and peptidolysis; NAS.
InterPro; IPR001314; Chymotrypsin.
InterPro; IPR002172; LDL_receptor_A.
InterPro; IPR001254; Ser_protease_Try.
InterPro; IPR001190; Srcr_receptor.
Pfam; PF00057; ldl_recept_a; 1.
Pfam; PF00083; trypsin; 1.
PRINTS; PR00722; CHYMOTRYPSIN.
SMART; SM00192; LDLa; 1.
SMART; SM00202; SR; 1.
SMART; SM00020; Tryp_SPC; 1.
PROSITE; PS01209; LDLRA_1; FALSE NEG.
PROSITE; PS50068; LDLRA_2; FALSE NEG.
PROSITE; PS00420; SRCR_1; FALSE NEG.
PROSITE; PS50287; SRCR_2; 1.
PROSITE; PS50240; TRYPSIN_DOM; 1.
PROSITE; PS00134; TRYPSIN_HIS; 1.
PROSITE; PS00135; TRYPSIN_SER; 1.
Hydrolase; Serine protease; Transmembrane; Signal-anchor..
DOMAIN 1 38
TRANSMEM 39 59
CYTOPLASMIC (POTENTIAL).
SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN
(POTENTIAL)).
EXTRACELLULAR (POTENTIAL).
LDL-RECEPTOR CLASS A.
SRCR.
SERINE PROTEASE.
CHARGE RELAY SYSTEM (BY SIMILARITY).
CHARGE RELAY SYSTEM (BY SIMILARITY).
CHARGE RELAY SYSTEM (BY SIMILARITY).
CLEAVAGE (POTENTIAL).
BY SIMILARITY.
BY SIMILARITY.
BY SIMILARITY.
BY SIMILARITY.
BY SIMILARITY.
BY SIMILARITY.
BY SIMILARITY.
N-LINKED (GLCNAC. .) (POTENTIAL).
N-LINKED (GLCNAC. .) (POTENTIAL).
MLQPDSDQPLNSLDVKEPKRPIMETFRK -> MSNPCA
NPVSEWRPSES (IN REF. 2).
NPVSEWRPSES (IN REF. 2).
351B2FD4A8657B12 CRC64;
SEQUENCE 437 AA; 48204 MW; 351B2FD4A8657B12 CRC64;
; Q9NRS4 Length: 437 November 25, 2003 13:16 Type: P Check: 316 ..
; q9nrs4
Query Match 79.1%; Score 1266; DB 1; Length 437;
Best Local Similarity 97.4%; Pred. No. 0;
Matches 228; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
QY 51 KIVGGYALDVSWPWQVSIQYDKQHVCGGSILDPHWLTAACHFRKHTDVFNNKVRAGSD 110
:|||||

```

```

Db 204 RVVGGEEASVDSWPWQVSIQYDKQHVCGGSILDPHWLTAACHFRKHTDVFNNKVRAGSD 263
QY 111 KLGSPFSLAVAKIIIEFNPMYPKNDIALMKLQPLTFSGTVRPICLFFDFDELTPTP 170
Db 264 KLGSPFSLAVAKIIIEFNPMYPKNDIALMKLQPLTFSGTVRPICLFFDFDELTPTP 323
QY 171 LMIIGWFTKQNGGKMSDILLQASVQVIDSTRCNADDAVQGEVTEKMCAGIPEGGVDTIC 230
Db 324 LMIIGWFTKQNGGKMSDILLQASVQVIDSTRCNADDAVQGEVTEKMCAGIPEGGVDTIC 383
QY 231 QGDSGGPLMYQSDQDQHVGVSVGWGCGGPGSTPGVVTKVSAVLNWIYNWKAE 284
Db 384 QGDSGGPLMYQSDQDQHVGVSVGWGCGGPGSTPGVVTKVSAVLNWIYNWKAE 437

```

Search completed: November 25, 2003, 13:51:12
Job time : 0.001 secs

GenCore version 5.1.6
 Copyright (c) 1993 - 2003 Compugen Ltd.
 OM protein - nucleic search, using frame_plus_p2n model
 Run on: November 25, 2003, 13:57:34 ; Search time 1.79505 Seconds
 (without alignments)
 2.016 Million cell updates/sec
 Title: US-09-607-745-2
 Perfect score: 2342
 Sequence: 1 MPPDSQPPLNSLDVKPLRP.....VTKVSAYLNWYVWKAEL 435

Scoring table: BLOSUM62
 Xgapop 10.0 , Xgapext 0.5
 Ygapop 10.0 , Ygapext 0.5
 Fgapop 6.0 , Fgapext 7.0
 Delop 6.0 , Delext 7.0

Searched: 2 seqs, 4160 residues
 Total number of hits satisfying chosen parameters: 4

Minimum DB seq length: 0
 Maximum DB seq length: 2000000000
 Post-processing: Minimum Match 0%
 Maximum Match 100%
 Listing first 45 summaries

Command line parameters:
 -MODE=frame+pn.model -DEV=soft -Q=US09607745.ppp -DB=nt.seq -SUFFIX=pto
 -OUT=nacomp.res -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0 -UNITS=bits -START=1
 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200
 -THR SCORE=ptc -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc
 -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000 -NCPU=6 -NO_XLPXY
 -NEG SCORES=0 -LONGLOG -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOF=6 -FGAPEXT=7
 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : nt.seq.*

Pred. No. is the number of results predicted by chance to have a
 score greater than or equal to the score of the result being printed,
 and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2337	99.8	2081	1 af179224	TOIG of: af17922
2	2334	99.2	2079	1 af216312	TOIG of: af21631
C 3	52.5	2.2	2081	1 af179224	TOIG of: af17922
C 4	46	2.0	2079	1 af216312	TOIG of: af21631

ALIGNMENTS

RESULT 1
 af179224
 ; TOIG of: af179224 check: 328 from: 1 to: 2081
 ; LOCUS AF179224 2081 bp mRNA linear PRI 08-JUN-2000
 ; DEFINITION Homo sapiens transmembrane serine protease 3 (TMPRSS3) mRNA,
 ; complete cds.
 ; ACCESSION AF179224
 ; VERSION AF179224.1 GI:8347148
 ; KEYWORDS
 ; SOURCE Homo sapiens (human)
 ; ORGANISM Homo sapiens
 ; Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 1 (bases 1 to 2081)
 Wallrapp,C., Rahmel,S., Muller-Pillasch,F., Burghardt,B.,
 Iwamura,T., Ruthenburger,M., Lerch,M.M., Adler,G. and Gress,T.M.
 A novel transmembrane serine protease (TMPRSS3) overexpressed in
 pancreatic cancer
 Cancer Res. 60 (10), 2602-2606 (2000)
 20283276
 10825129
 2 (bases 1 to 2081)
 Wallrapp,C. and Gress,T.M.
 Direct Submission
 Submitted (20-AUG-1999) Internal Medicine I, University of Ulm,
 Robert-Koch-Street 8, Ulm, Baden-Wuerttemberg 89081, Germany
 Location/Qualifiers
 1. 2081
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /chromosome="11"
 /map="11q23.3"
 /tissue_type="pancreatic carcinoma"
 1. 2081
 /gene="TMPRSS3"
 215..1528
 /gene="TMPRSS3"
 /codon_start=1
 /product="transmembrane serine protease 3"
 /protein_id="AAP74526.1"
 /db_xref="GI:8347148"
 /translation="MLQDPDSQPLNSLDVKPLRPIMETPRKVGIPILALLSLA
 AVLSKDRSTQLVDSATGNWAFEDNFTALAEATACRQGYSSKPTFAVEIGPQV
 DLVVEITENSQELRMNSGCLSGSLVSLHCLACGSLKTRVVGGEASVDSWV
 QVSIQDKOHCVCGSILDPHWLTAHCRKHTDFVFNKVRAGSDKLGSPSLAVAKI
 IIEFNPYKDNIDILMKIQPLTFSTGVRPCLPFDESLTPATPLIIGWFTQ
 NGKMSDILLOASVQVLDSTRCNADDAQCQVTEKMCAGIPGGVDTCCGDSGGLM
 YQSDQHVGVGVSWGCGGPGSTPGVTKVSAIYNWVWKAEL"
 215..307
 /gene="TMPRSS3"
 /note="Region: cytoplasmic domain"
 308..373
 /gene="TMPRSS3"
 /note="Region: transmembrane domain"
 374..1525
 /gene="TMPRSS3"
 /note="Region: extracellular domain"
 602..604
 /gene="TMPRSS3"
 /note="Glycosylation site"
 746..748
 /gene="TMPRSS3"
 /note="glycosylation site"
 order(800..802,1142..1144)
 /gene="TMPRSS3"
 /note="disulfide bond"
 order(902..904,950..952)
 /gene="TMPRSS3"
 /note="disulfide bond"
 947..949
 /gene="TMPRSS3"
 /note="active site"
 1082..1084
 /gene="TMPRSS3"
 /note="active site"
 order(1280..1282,1328..1330)
 /gene="TMPRSS3"
 /note="disulfide bond"
 1373..1375
 /gene="TMPRSS3"
 /note="active site"
 484 a 597 c 576 g 424 t
 BASE COUNT
 ORIGIN

AF179224 Length: 2081 November 25, 2003 13:15 Type: N Check: 328 ..
af179224

Alignment Scores:

Pred. No.: 0 Length: 2081
Score: 2337.00 Matches: 434
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 99.79% Indels: 0
DB: 1 Gaps: 0

US-09-607-745-2 (1-435) x af179224 (1-2081)

```

QY 2 AspProAspSerAspGlnProLeuAsnSerLeuAspVallyysProLeuArgLysProArg 21
DB 224 GATCTCGACAGTATCAACCTCTGACAGCCTCGATGTCACAAACCCCTGCGCAACCCCGT 283
QY 22 IleProMetGluThrPheArgLysValGlyIleProIleIleLeuLeuSerLeu 41
DB 284 ATCCCATGGAGACCTTCAGAAAGGTGGGATCCCATCATATAGCAGTACTGAGCCTG 343
QY 42 AlaSerIleIleValValValValLeuLeuLeuVallyysVallyysVallyysPheLeu 61
DB 344 GCGAGTATCATATTGTTGTTCTCATCAAGGTGATTCGGATAAATACTACTTCTCTC 403
QY 62 CysGlyGlnProLeuHisPheIleProArgLysGlnLeuCysAspGlyGluLeuAspCys 81
DB 404 TCGGGGCGACCTCTCATCTTATCCCGAGGAGAGCTGTGTGACGAGAGCTGAGCTGT 463
QY 82 ProLeuGlyGluAspGluHisCysVallyysSerPheProGluGlyProAlaValAla 101
DB 464 CCCTTGGGGAGGACGAGGAGCAGTGTGTCAAGAGCTTCCCGAAGGCGCTGCGAGTGC 523
QY 102 ValArgLeuSerLysAspArgSerThrLeuGlnValLeuAspSerAlaThrGlyAsnTrp 121
DB 524 GTCCCGCTCTCCAGAGCAGGATCCACATCGACATGCGAGGTGTGGACTCGGCCACAGGA 583
QY 122 PheSerAlaCysPheAspAsnPheThrGluAlaLeuAlaGluThrAlaCysArgGlnMet 141
DB 584 TTCTCTGCTGCTTTTCGACAACTTCACAGAGCTCTCGCTGAGACAGCTGTAGGCAGATG 643
QY 142 GlyTyrSerSerLysProThrPheArgAlaValGluIleGlyProAspGlnAspLeuAsp 161
DB 644 GGCTACAGCAGCAAAACCCACTTTCAGAGTGTGGAGATGGCCCAAGACAGCAGATCTGG 703
QY 162 ValValGluIleThrGluAsnSerGlnGluLeuArgMetArgAsnSerSerGlyProCys 181
DB 704 GTTGTGGAAATACAGAAACAGCCAGAGCTTCGATGGGAACTCAAGTGGGCCCTGT 763
QY 182 LeuSerGlySerLeuValSerLeuHisCysLeuAlaCysGlyLysSerLeuLysThrPro 201
DB 764 CTCTCAGGCTCCCTGCTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 823
QY 202 ArgValValGlyGlyGluGluAlaSerValAspSerTrpProTrpGlnValSerIleGln 221
DB 824 CGTGTGGTGGTGGGAGAGGAGGCTCTGAGATCTTGGCCCTTGGCAGGTGAGATCCAG 883
QY 222 TyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThrAla 241
DB 884 TACGACAAACAGCAGCAGCTCTGTGGAGGAGCATCTCGACCCCACTGGGTCTCCACGG 943
QY 242 AlaHisCysPheArgLysThrAspValPheAsnTrpLysValArgAlaGlySerAsp 261
DB 944 GCCCATGCTTTCAGAAACATACCGATGTTCATCGAAGAGGTGGGAGGCTCAGAC 1003
QY 262 LysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleGluPheAsnPro 281
DB 1004 AAATGGGAGAGCTTCCATCCCTGGCTGTGGCCAGATCATATATTGAATCAACCCC 1063
QY 282 MetTyrProLysAspAsnAlaLeuMetLysLeuGlnPheProLeuThrPheSer 301
DB 1064 ATGTACCCCAAGACAATGACATCGCCCTCATGAGCTGAGTTCCTCACTTCTCTCA 1123

```

```

QY 302 GlyThrValArgProIleCysLeuProPheAspGluLeuThrProAlaThrPro 321
DB 1124 GGCACAGTCAGGCCCATCTGTCGCCCTTCTTTGATGAGGAGCTCCTCCAGGCCACCCCA 1183
QY 322 LeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyGlyLysMetSerAspIleLeu 341
DB 1184 CTCTGGATCANTGATGGGCTTTACGAGCAGATGAGGAGGAGATGCTGACATCTG 1243
QY 342 LeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAspAlaTrpGln 361
DB 1244 CTGACGGGCTCAGTCCAGGTCAATGACACACACGGTCAATGACAGCAGTGCCTACAG 1303
QY 362 GlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThrCys 381
DB 1304 GGGGAGTCAACGAGAGATGATGTGTCAGGATCCCGGAGGGGTGTGGACACCTGC 1363
QY 382 GlnGlyAspSerGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGlyIle 401
DB 1364 CAGGGTGCAGTGTGGGCCCTGATGTACCAATCTGACCAATGTCATGTGTGGGCATC 1423
QY 402 ValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyValTrpThrLysValSer 421
DB 1424 GTTACTGGGGATGATGGTGTGGGGGCCCGGAGCACCACCGAGGTATACACCAAGTCTCA 1483
QY 422 AlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu 435
DB 1484 GCCTATCTCACTGGATCTACAATGCTGTGGAAGGCTGAGCTG 1525

```

RESULT 2
af216312
TOIG of: af216312 check: 3320 from: 1 to: 2079

LOCUS AF216312 2079 bp mRNA linear PRI 07-FEB-2000
DEFINITION Homo sapiens type II membrane serine protease mRNA, complete cds.
ACCESSION AF216312
VERSION AF216312.1 GI:6911218
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
MAMMALIA; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 2079)
Smeekens,S.S., Lorimer,D.D., Wang,E., Hou,J. and Linnevers,C.
MT-SP2, a novel type II membrane serine protease expressed in
trachea, colon, and small intestine: identification, cloning, and
chromosomal localization
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2079)
Smeekens,S.S., Lorimer,D.D., Wang,E., Hou,J. and Linnevers,C.
Direct Submission
TITLE Submitted (14-DEC-1999) Axy's Pharmaceuticals, Inc, 180 Kimball Way,
South San Francisco, CA 94080, USA
JOURNAL Location/Qualifiers
FEATURES
1. 2079
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
251..1522
/note="MT-SP2"
/codon_start=1
/product="type II membrane serine protease"
/protein_id="AA031436.1"
/db_xref="GI:6911218"
/translation="MSNFCANFVSPWPSVGVPIIIALLSLASIIIVVVLKVLID
KYFLCOQLHFFIPKQLCDGLDPLGDEEHCKVSFFPEGPAVALRSKDRSTLQVL
DSATGMFSAFDFNTEALACQMGYSKPTFRAVEIPDQDLVDLVEITENSDEL
RMNRSGLCSGLSLHSLACQKSLKTPRWGVEASVDSWPQVSIQDKQHVCGG
SILDHPVLTAHCFRKHDTDVNWKVRAGSKLGSFSLAVAKIIIEFNPMYKND
IALMKLQPLTFSTGTPICLPFDEBELTATPLIILGWGFTKQNGSKSDILLQASV
QVIDSTCNADDAVQGEVTEKMCAGLPEGGVDTCQDSCGGLMYQSDQHVVGIVSW
GYCGGFSFTGVTKVSNLWYNVWKAEL"
BASE COUNT 489 a 594 c 575 g 421 t

```

; ORIGIN.
; AF216312 Length: 2079 November 25, 2003 13:16 Type: N Check: 3320 ..
af216312
Alignment Scores:
Pred. No.: 0 Length: 2079
Score: 2324.00 Matches: 434
Percent Similarity: 99.77% Conservative: 0
Best Local Similarity: 99.77% Mismatches: 0
Query Match: 99.23% Indels: 1
DB: 1 Gaps: 0

US-09-607-745-2 (1-435) x af216312 (1-2079)
QY 2 AspProAspSerAspGlnProLeuAsnSerLeuAspValLysProLeuArgLysProArg 21
Db 217 GATCCTGACAGTGATCAACCTCTGAACAGAGCTCGATGTCACAAACCCCTGGCGAACCCCGT 276
QY 22 IleProMetGluThrPheArgLys-ValGlyIleProIleIleIleIleLeuSerLe 41
Db 277 ATCCCCATGAGACCTTCAGAAAGTGTGGGATCCCCATCATCATGACACTATCTGAGCCT 336
QY 41 uAlaSerIleIleIleValValValLeuIleLysValIleLeuAspLysTyrTyrPheLe 61
Db 337 GCGCAGTATCATATGTGTGTTGTCCTCATCAAGGTGATCTGGATTAATPACTATCTTCC 396
QY 61 uCysGlyGlnProLeuHisPheLeProArgLysGlnLeuCysAspGlyLeuAspCy 81
Db 397 CTGCGGCGACCTCTCCACTTCATCCCGAGAGACAGCTGTGTGACGAGAGCTGGACTG 456
QY 81 sProLeuGlyGluAspGluCysValLysSerPheProGluGlyProIleValAla 101
Db 457 TCCCTTGGGGAGAGAGAGAGAGACTGTGTCAAGAGCTTCCCGAAGGGCTCGAGTGC 516
QY 101 aValArgLeuSerLysAspArgSerThrLeuGlnValLeuAspSerAlaThrGlyAsnTr 121
Db 517 AGTCGCCCTCTCCAGGACCGATCCACACTGCAGTGTGTGATCTGGCCACAGGAACTG 576
QY 121 pPheSerAlaCysPheAspAsnPheThrGluAlaLeuAlaGluThrAlaCysArgGlnMe 141
Db 577 GTTCTCTGCCCTTTTCACAACTTCACAGAACTCTCGCTGAGACAGCCTGTAGGCAGAT 636
QY 141 tGlyTyrSerSerLysProThrPheArgAlaValGluIleGlyProAspGlnAspLeuAs 161
Db 637 GGCTACAGCAGCAAAACCCACTTTCAGAGCTGTGGAGATTGGCCACAGGATCTGGA 696
QY 161 pValValGluIleThrGluAsnSerGlnGluLeuArgMetArgAsnSerSerGlyProCy 181
Db 697 TGTGTGTGAATCACAGAAACAGCCAGAGCTTCGCATCCGGAACCTCAAGTGGGCCCTG 756
QY 181 sLeuSerGlySerLeuValSerLeuHisCysLeuAlaCysGlyLysSerLeuLysThrPr 201
Db 757 TCTCTCAGGCTCCCTGCTCTCCCTGCACTCTCTGCTGTGGAGAGAGCTGAAGACCCC 816
QY 201 oArgValValGlyGlyGluIleAserValAspSerTrpProGlnValSerIleG 221
Db 817 CCGTGTGGTGGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 876
QY 221 nTyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThrAl 241
Db 877 GTACGACAAACACACAGCTCTGTGGAGGAGAGATCTTGGACCCCACTTGGGTCTCTACGGC 936
QY 241 aAlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySerAs 261
Db 937 AGCCCACTGCTTCAGGAAACATACCGATGTGTTCAACTGGAAGTGGCGGAGGCTCAGA 996
QY 261 pLysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleGluPheAsnPr 281
Db 997 CAAACTGGGAGAGTCTCCATCCCTGGCTGTGGCGCAAGATCATCATGAAATCAACCC 1056
QY 281 oMetTyrProLysAspAspIleAlaLeuMetLysLeuGlnPheProLeuThrPheSe 301

```

RESULT 3

af179224/c

; TOIG of: af179224 check: 328 from: 1 to: 2081

```

; LOCUS AF179224 2081 bp mRNA linear PRI 08-JUN-2000
; DEFINITION Homo sapiens transmembrane serine protease 3 (TMPRSS3) mRNA,
; complete cds.
; ACCESSION AF179224
; VERSION AF179224.1 GI:8347148
; KEYWORDS .
; SOURCE Homo sapiens (human)
; ORGANISM Homo sapiens
; Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
; REFERENCE 1 (bases 1 to 2081)
; AUTHORS Wallrapp,C., Hahnel,S., Muller-Pillasch,F., Burghardt,B.,
; Iwamura,T., Ruthenburger,M., Leich,M.M., Adler,G. and Gress,T.M.
; A novel transmembrane serine protease (TMPRSS3) overexpressed in
; pancreatic cancer
; JOURNAL Cancer Res. 60 (10), 2602-2606 (2000)
; MEDLINE 20283276
; PUBMED 10825129
; REFERENCE 2 (bases 1 to 2081)
; AUTHORS Wallrapp,C. and Gress,T.M.
; TITLE Direct Submission
; JOURNAL Submitted (20-AUG-1999) Internal Medicine I, University of Ulm,
; Robert-Koch-Street 8, Ulm, Baden-Wuerttemberg 89081, Germany
; FEATURES
; Location/Qualifiers
; 1..2081
; /organism="Homo sapiens"
; /mol_type="mRNA"
; /db_xref="taxon:9606"
; /chromosome="11"
; /map="11q23.3"
; /tissue_type="pancreatic carcinoma"
; 1..2081
; /gene="TMPRSS3"
; 215..1528
; /gene="TMPRSS3"
; /codon_start=1
; /product="transmembrane serine protease 3"
; /protein_id="AAF74526.1"
; /db_xref="GI:8347149"

```

```

/translation="MLQPDSDQPLNSLDVKKPRIPMETFRKVGIPILIALLSLA
SIIVVILKIDKYLFCGQPLHPIRKQCDGELDCPLGDEEHCVKSPFPEPAV
AVRLSKRSTLQVLDASATNWFSAFDFNTEALAEATACROMGYSSKPTFRVETGPDQ
DLVVVITENSOELMRNSGFCLSGLSVLSHLACAGKSLKTPRVVGGEEASVDSWPW
QVSIQYQKHVCGGSLDPHWLTAHCFRKHDTVFNKVRAGSKLGFPPSLAVAKI
IIBPNMPYKGDNDIALMKLOPLPFTSGTVRPICLPFFDEELTPATPLWIIIGWFTKO
NGKMSDILLQASVOIDSTRCNADDAVCGEVTOMCAGIPEGVDTCQSDSGGFLM
YQSDQHWVIGVSWGCGGPGTPGVYTKVSAYLNIWVNWKAEL"
215..307
/misc_feature
/gene="TMPRSS3"
/note="Region: cytoplasmic domain"
308..373
/misc_feature
/gene="TMPRSS3"
/note="Region: transmembrane domain"
374..1525
/misc_feature
/gene="TMPRSS3"
/note="Region: extracellular domain"
602..604
/misc_feature
/gene="TMPRSS3"
/note="glycosylation site"
746..748
/misc_feature
/gene="TMPRSS3"
/note="glycosylation site"
order(800..802,1142..1144)
/misc_feature
/gene="TMPRSS3"
/note="disulfide bond"
order(902..904,950..952)
/misc_feature
/gene="TMPRSS3"
/note="disulfide bond"
947..949
/misc_feature
/gene="TMPRSS3"
/note="active site"
1082..1084
/misc_feature
/gene="TMPRSS3"
/note="active site"
order(1280..1282,1328..1330)
/misc_feature
/gene="TMPRSS3"
/note="disulfide bond"
1373..1375
/misc_feature
/gene="TMPRSS3"
/note="active site"
484 a 597 c 576 g 424 t
BASE COUNT
ORIGIN
AF179224 Length: 2081 November 25, 2003 13:15 Type: N Check: 328 ..
af179224
Alignment Scores:
Pred. No.: 0 Length: 2081
Score: 52.50 Matches: 52
Percent Similarity: 35.62% Conservative: 26
Best Local Similarity: 23.74% Mismatches: 86
Query Match: 2.24% Indels: 57
DB: 1 Gaps: 10
US-09-607-745-2 (1-435) x af179224 (1-2081)
QY 23 ProMetGluThrPheArgLysValGlyLe-----ProIleIleLeaLeuLeu 39
Db 684 CCAATCTCCACAGCTCTGAAGAGGGTTCCTGCTGTAGCCCTCTACAGGCTGTC 625
QY 40 SerLeuAlaSerIleIleLeValValLeuLeuLeuValIleLeuAspLysTyr--- 58
Db 624 TCAGCAGAGCT-----TCTGTGAAGTTCGGAACAGGCA 589
QY 59 -----Ty-PheLeuCysGlyGlnProLeuHisPheIleProArgLysGlnLeuCysAsp 76
Db 588 GAGAACCAAGTTCCC-TGTGGCCGAGCCACCACTGCAGTGGATCGGTCTT----- 536
QY 77 GlyGluLeuAspCys-----ProLeuGlyGluAspGluGluHisCysValLysSer 93
Db 535 GGAGAGCGGAGCTGCCACTGCAGCCCTTCGGGGAAGCTCTTGACACAGTCTCCTCGTC 476

```

```

QY 94 PheProGluGlyProAla---ValAlaValArgLeuSerLysAspArgSerThrLeuGln 112
Db 475 CTCCCCAAGGAGCAGTCCAGCTCTCCGTCACACAGCTGCTTCTCGGGATGAAGTGGAG 416
QY 113 ValLeu-----AspSerAlaThr 118
Db 415 AGCTCCCGCAGAGAGTAGTATTATCCAGATCATCCTTGATGAGGACCAACCAAT 356
QY 119 GlyAsnTrpPheSerAla-----CysPheAspAsnPheThrGluAlaLeuAlaGluThr 136
Db 355 GATGATACTCGCCAGAGCTCAGTAGTCTATGATGATGGGATCCCCACCTTTCTGAGGT 296
QY 137 AlaCysArgGlnMetGlyTyrSerSerLysProThrPheArgAlaValGluLeuGlyPro 156
Db 295 CTCCATGGGGATACGGGGTTTGGCAGAGGGGTTTGACATCGAGGCTGTT----- 248
QY 157 AspGlnAspLeuAspValValGluLeuThrGluAsnSerGlnGluLeuArgMetArgAsn 176
Db 247 ---CAGAGGTTGATCACTGTGAGGATCTGTAAAT----- 215
QY 177 SerSerGlyProCysLeuSerGlySerLeuValSerLeuHisCysLeuAlaCysGlyLys 196
Db 214 -----GCTGGCTCTGTGATCTCTCCGGTCTCCCTGTAGCTGGAGCT 173
QY 197 SerLeuLysThrProArgValValGlyGlyGluGluAlaSerValAspSerTrpPro 215
Db 172 GAGATTGTCAACCCCAAGGCA-GCAGGAGGGCTCCCCACACAGGCTCTGCTGGCCA 117
RESULT 4
af216312/c
; TOIG of: af216312 check: 3320 from: 1 to: 2079
; LOCUS AF216312 2079 bp mRNA linear PRI 07-FEB-2000
; DEFINITION Homo sapiens type II membrane serine protease mRNA, complete cds.
; ACCESSION AF216312
; VERSION AF216312.1 GI:6911218
; KEYWORDS
; SOURCE Homo sapiens (human)
; ORGANISM Homo sapiens
; Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
; REFERENCE 1 (bases 1 to 2079)
; AUTHORS Smeekens,S.S., Lorimer,D.D., Wang,E., Hou,J. and Linnevers,C.
; TITLE MT-S2, a novel type II membrane serine protease expressed in
; trachea, colon, and small intestine: identification, cloning, and
; chromosomal localization
; JOURNAL Unpublished
; REFERENCE 2 (bases 1 to 2079)
; AUTHORS Smeekens,S.S., Lorimer,D.D., Wang,E., Hou,J. and Linnevers,C.
; TITLE Direct Submission
; JOURNAL Submitted (14-DEC-1999) Axys Pharmaceuticals, Inc, 180 Kimball Way,
; South San Francisco, CA 94080, USA
; FEATURES
; 1..2079
; Location/Qualifiers
; /organism="Homo sapiens"
; /mol_type="mRNA"
; /db_xref="taxon:9606"
; 251..1522
; /notes="MT-SP2"
; /codon_start=1
; /product="type II membrane serine protease"
; /protein_id="AAF31436.1"
; /db_xref="GI:6911219"
; /translation="MSNPANFVSPWRPSEVGIPIIIALLSLAIIIVVILKVID
KYVLCGQPLHPIRKQCDGELDCPLGDEEHCVKSPFPEPAVAVRLSKRSTLQVL
DSATGNWFSACDNFTFALAEATACROMGYSSKPTFRVETGPDQDLVDVETNSREL
RMRNSGFCLSGLSVLSHLACAGKSLKTPRVVGGEEASVDSWPWQVSIQYQKHVCGG
SLDPHWLTAHCFRKHDTVFNKVRAGSKLGFPPSLAVAKI IIBPNMPYKGDND
IALMKLOPLPFTSGTVRPICLPFFDEELTPATPLWIIIGWFTKONGKMSDILLQASV
OVIDSTRCNADDAVCGEVTOMCAGIPEGVDTCQSDSGGFLMYQSDQHWVIGVSW
GYGCGSPSTPGVYTKVSAYLNIWVNWKAEL"
; BASE COUNT 489 a 594 c 575 g 421 t
; ORIGIN

```


Maximum DB seq length: 200000000

```

;
; OVIDTRCNADDAAYOEYTERKMWKAGIPRGGVDTTCQSDSGGFLMYQSDQWHVGVISW
; GYGGCGSPSPGYTKVSAYLNIYNWKAEI"
; BASE COUNT      489 a      594 c      575 g      421 t
; ORIGIN
;
; AF179224 Length: 2079 November 25, 2003 13:16 Type: N Check: 3320 ..
af216312
Alignment Scores:
Pred. No.:      0      Length:      2079
Score:      1266.00      Matches:      228
Percent Similarity:      98.29%      Conservative:      2
Best Local Similarity:      97.44%      Mismatches:      4
Query Match:      79.12%      Indels:      0
DB:      1      Gaps:      0

US-09-607-745-9 (1-292) x af216312 (1-2079)
QY 51 LysIleValGlyGlyTyrAlaLeuAspValAspSerTrpProTrpGlnValSerIleGln 70
Db 818 CGTGTGCTGGTGGGAGGAGCCCTCTGTGGATCTTGGCTTGGCAGGTGACATCCAG 877
QY 71 TyrAspLysGlnHisValCysGlySerIleLeuAspProHisTrpValLeuThrAla 90
Db 878 TAGCACAAACAGCAGCTCTGTGGAGGAGCATCTCGACCCCACTGGTCTCCTCAGGCA 937
QY 91 AlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySerAsp 110
Db 938 GCCACTGCTTCAGAAACATACCATGTGTCTCACTGGAGGTGGCGGAGGCTCAGAC 997
QY 111 LysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleGluPheAsnPro 130
Db 998 AAACCTGGCAGCTCCATCCCTGGCTGTGGCAGATCATCATCATGATTAATCAACCCC 1057
QY 131 MetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPheSer 150
Db 1058 ATGTACCCCAAGACAAATAGATGATCGCCCTCATGAAGCTGAGTTCACACTCTCTCA 1117
QY 151 GlyThrValArgProIleCysLeuProPhePheAspGluGluLeuThrProAlaThrPro 170
Db 1118 GGCACAGTCAGGCCCATCTCTGTGCCCTTCTTGTATGAGGAGCTCATCCAGCACCCCA 1177
QY 171 LeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyGlyLysMetSerAspIleLeu 190
Db 1178 CTCTGGATCATTTGATGGGCTTTACGAGCAGATGAGGAGAGATGCTGTGACACTCTG 1237
QY 191 LeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAlaTyrGln 210
Db 1238 CTCGAGCGTCAGTCCAGGTCATGTGACAGCACACGCGTGCATGCGACGATGGTACCAG 1297
QY 211 GlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyValAspThrCys 230
Db 1298 GGGGAAGTCACCGAGAGATGATGTGCGAGGATCCCGAAGGGGTGGACACTGCTC 1357
QY 231 GlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValGlyIle 250
Db 1358 CAGGGTGACAGTGTGGGCCCTTGATGATCAATCTGACCAATGCGATGTGGTGGCATC 1417
QY 251 ValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyValTyrThrLysValSer 270
Db 1418 GTTAGCTGGGGCTATGGCTCGGGGGGCGCCAGCACCCCGAGGATATACCAAGGTCTCA 1477
QY 271 AlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu 284
Db 1478 GCCTATCTCACTGGATCTACATATGCTGAGAGGCTGAGCTG 1519

RESULT 2
af179224
; TOIG of: af179224 check: 328 from: 1 to: 2081
;
; LOCUS      AF179224      2081 bp      mRNA      linear      PRI 08-JUN-2000
; DEFINITION      Homo sapiens transmembrane serine protease 3 (TMPRSS3) mRNA,
; complete cds.

```

```

; ACCESSION      AF179224
; VERSION        AF179224.1      GI:8347148
; KEYWORDS
; SOURCE
; ORGANISM       Homo sapiens (human)
;
; Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
;
; 1 (bases 1 to 2081)
; AUTHORS       Wallrapp,C., Hahnel,S., Muller-Pillasch,F., Burghardt,B.,
;               Iwamura,T., Ruthenburger,M., Lerch,M.M., Adler,G. and Gress,T.M.
; TITLE        A novel transmembrane serine protease (TMPRSS3) overexpressed in
;               pancreatic cancer
; JOURNAL       Cancer Res. 60 (10), 2602-2606 (2000)
; MEDLINE       20283276
; PUBMED        10825129
; REFERENCE     2 (bases 1 to 2081)
;               Wallrapp,C. and Gress,T.M.
; TITLE        Direct Submission
; JOURNAL       Submitted (20-AUG-1999) Internal Medicine I, University of Ulm,
;               Robert-Koch-Street 8, Ulm, Baden-Wuerttemberg 89081, Germany
; FEATURES
;   source
;     1..2081
;       /organism="Homo sapiens"
;       /mol_type="mRNA"
;       /db_xref="taxon:9606"
;       /chromosome="11"
;       /map="11q23.3"
;       /tissue_type="pancreatic carcinoma"
;     1..2081
;       /gene="TMPRSS3"
;       /cds
;         215..1528
;         /gene="TMPRSS3"
;         /product="transmembrane serine protease 3"
;         /protein_id="AAF74526.1"
;         /db_xref="GI:8347148"
;         /translation="MLQDDPSDDPLNSLDVPLKPKRIPMETRKVGIPILIALSLA
;         SILIIVLLIKVILDKYIFLCQLHFIPIRKQLCGELDCPLGDEEHCVKSFPPGPAP
;         AVLRSDKSTLQVDSAINGNWFSCDFNFTEALAEATACRMQYSSKPTFAVEIGPDQ
;         DLDVETITENSQELMRNNSGCLSGSLVSLHCLACGLKLTPTRVGGEASVDSWPW
;         QVSIQDKQHVCGSILDPHWLTAAHCFKRTDVFNVKVRAGSKLGSFSLAVAKI
;         IITEFPMYKDNIDIALMKLQPLPTPSGTVPRTCLPFDEBELTPATPLIILGWGFTKQ
;         NGKMSDILLQASVQVISTRCNADADAYQSEVTEKMKACGIPGGGVDTTCQSDSGGFLM
;         YQSDQHVGVGVISWGGGSPFGVITTKVSAYLNIYNWKAEI"
;       215..307
;       /gene="TMPRSS3"
;       /notes="Region: cytoplasmic domain"
;       308..373
;       /gene="TMPRSS3"
;       /note="Region: transmembrane domain"
;       374..1525
;       /gene="TMPRSS3"
;       /notes="Region: extracellular domain"
;       602..604
;       /gene="TMPRSS3"
;       /note="glycosylation site"
;       746..748
;       /gene="TMPRSS3"
;       /note="glycosylation site"
;       /order(800..802,1142..1144)
;       /gene="TMPRSS3"
;       /note="disulfide bond"
;       /order(902..904,950..952)
;       /gene="TMPRSS3"
;       /note="disulfide bond"
;       947..949
;       /gene="TMPRSS3"
;       /note="active site"
;       1082..1084
;       /gene="TMPRSS3"
;       /note="active site"
;       /order(1280..1282,1328..1330)
;       /gene="TMPRSS3"

```



```

QY 224 -----GlyValAspThrCysGlnGly----- 232
Db 1567 GGGCAGGAGGAGCGGCTCCACGACTGCAAGGGGCGAGCAGCATTCACGCTCAGCCTT 1508
QY 233 ---AspSerGlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGlyVal 251
Db 1507 CCAGACATTGTAGATCCAGTGTAGATAGCTGAGAC-----CTGTGTGTATAC 1460
QY 252 SerTrpGly 254
Db 1459 TCCTGGGGT 1451

RESULT 4
af179224/c
; TOIG of: af179224 check: 328 from: 1 to: 2081
; LOCUS AF179224 2081 bp mRNA linear PRI 08-JUN-2000
; DEFINITION Homo sapiens transmembrane serine protease 3 (TMPRSS3) mRNA,
; complete cds.
; ACCESSION AF179224
; VERSION AF179224.1 GI:8347148
; KEYWORDS
; SOURCE Homo sapiens (human)
; ORGANISM Homo sapiens
; Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
; REFERENCE 1 (bases 1 to 2081)
; AUTHORS Wallrapp,C., Hannel,S., Muller-Pillasch,F., Burghardt,B.,
; Iwamura,T., Ruthenburger,M., Lerch,M.M., Adler,G. and Gress,T.M.
; TITLE A novel transmembrane serine protease (TMPRSS3) overexpressed in
; pancreatic cancer
; JOURNAL Cancer Res. 60 (10), 2602-2606 (2000)
; MEDLINE 20283276
; PUBMED 10825129
; REFERENCE 2 (bases 1 to 2081)
; AUTHORS Wallrapp,C. and Gress,T.M.
; TITLE Direct Submission
; JOURNAL Submitted (20-AUG-1999) Internal Medicine I, University of Ulm,
; Robert-Koch-Street 8, Ulm, Baden-Wuerttemberg 89081, Germany
; FEATURES
; source
; 1..2081
; /organism="Homo sapiens"
; /mol_type="mRNA"
; /db_xref="taxon:9606"
; /chromosome="11"
; /map="1q23.3"
; /tissue_type="pancreatic carcinoma"
; 1..2081
; /gene="TMPRSS3"
; 215..1528
; /gene="TMPRSS3"
; /codon_start=1
; /product="transmembrane serine protease 3"
; /protein_id="AAF74526.1"
; /db_xref="GI:8347149"
; translation="MLQDPDSQPLNSLDVKLPKPRIPMETEKVGIPILIALLSLA
; SIITVVLIVKLDKVFELGQPLHPIPRQLCDGELDCPLGEDBECVKSFPFGPAV
; AVRLSKRSTLOVLDSITGNWFSACFNFTALARTACRMGYSSKFTFRVAVIGPDQ
; QDVVEITENSQELRNNSGFCLSGLSLVHLCLACSKLTKPRVGGSEASVDSNFW
; QLSYQDKQVCGGSLIDPHWLTAHCFKRKHTDVFNVKVRAGSDKLGSPFSLAVAKI
; IIEFPMYPKNDIALMKLQPLTFSGVRPICLPFFDELTPATPLWIGWFTKQ
; NGKMSDILLQASQVVIDSRCDADAVQGVTEKMCAGIPEGGVDTCCQDSSGFLM
; YQSDQHVVGIVSVNGVCGGSPSTGVTVKVSAILNWNINWKAEI"
; 215..307
; /misc_feature
; /gene="TMPRSS3"
; /note="Region: cytoplasmic domain"
; 308..373
; /misc_feature
; /gene="TMPRSS3"
; /note="Region: transmembrane domain"
; 374..1525
; /misc_feature
; /gene="TMPRSS3"
; /note="Region: extracellular domain"

```

```

; misc_feature 602..604
; /gene="TMPRSS3"
; /note="glycosylation site"
; 746..748
; misc_feature
; /gene="TMPRSS3"
; /note="glycosylation site"
; order(800..802,1142..1144)
; /gene="TMPRSS3"
; /note="disulfide bond"
; order(902..904,950..952)
; /gene="TMPRSS3"
; /note="disulfide bond"
; 947..949
; misc_feature
; /gene="TMPRSS3"
; /note="active site"
; 1082..1084
; misc_feature
; /gene="TMPRSS3"
; /note="active site"
; order(1280..1282,1328..1330)
; /gene="TMPRSS3"
; /note="disulfide bond"
; 1373..1375
; misc_feature
; /gene="TMPRSS3"
; /note="active site"
; 484 a 597 c 576 g 424 t
; BASE COUNT
; ORIGIN
; AF179224 Length: 2081 November 25, 2003 13:15 Type: N Check: 328 ..
af179224
Alignment Scores:
Pred. No.: 0 Length: 2081
Score: 37.50 Matches: 36
Percent Similarity: 33.33% Conservative: 14
Best Local Similarity: 24.00% Mismatches: 55
Query Match: 2.34% Indels: 46
DB: 1 Gaps: 8
US-09-607-745-9 (1-292) x af179224 (1-2081)
QY 138 IleAlaLeuMetLysLeuGlnPheProLeuThrPheSerGlyThrValArgProIleCys 157
Db 1952 CTGTGCTAGTAGGCTTGGGGATGGTGAAACGACAGCGGCTGGCGACACCC----- 1899
QY 158 LeuProPhePheAspGluGluLeuThrProAlaThrProLeuTrpIleIleGlyTrpGly 177
Db 1898 -----TTTCCTTCTCCTCTCCAGCCACCA-----GTGATCTGGGCT 1863
QY 178 PheThrLysGlnAsnGlyGlyLysMetSerAspIleLeuGlnAlaSerValGln--- 196
Db 1862 TTACAGACAG-----CCTGCTTCCATTACAGTAG 1833
QY 197 -----ValIleAspSerThrArgCysAsnAlaAspAspAlaTyrGlnGlyGluValThr 214
Db 1832 TGTGGGAAGTTCTTCTTGGCTTAGCATATACCCCTGAGACCTTGTTCAGTGGGCTGT 1773
QY 215 GluLysMetMetCysAlaGlyLePro-----GluGlyGlyValAspThrCysGln 231
Db 1772 CTCTCCCTGGGATGCTG-GGAGCACCAGTGTGGCGGAGCTAGGCTGTGCTTCTCTCT 1714
QY 232 Gly-----AspSerGlyGlyProLeuMetTyrGlnSer 242
Db 1713 GGGCCCTCTCTGGGTGGCGAGGTCTCTTATAGGAATTCAGGCCCTTCTGCTCCAGAA 1654
QY 243 AspGlnTrpHisValValGlyLeValSerTrpGlyTyrGlyCysGlyGlyProSerThr 262
Db 1653 -----ATGCTGAGGCTGTGGCAGAGGGGTGT-----ACC 1624
QY 263 ProGlyValTyrThrLysValSerAlaTyr 272
Db 1623 CAAGGGGACTCTGTCTGTCTGTCTGACTTT 1594

```

Tue Nov 25 13:58:05 2003

nacompare.res

Page 9

Search completed: November 25, 2003, 13:57:43
Job time : 3.20495 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 23, 2003, 07:38:25 ; Search time 64.0234 Seconds
(without alignments)
1078.451 Million cell updates/sec

Title: US-09-607-745-2

Perfect score: 2342

Sequence: 1 MPDSDQPLNSLDVRLKRP.....VYTKSAYLNWYVWKAEL 435

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A Geneseq 19Jun03.*

1: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	2342	100.0	435	22	Human seripancrin
2	2342	100.0	435	23	Human serine protease
3	2342	100.0	435	24	Human expressed pr
4	2342	100.0	435	24	Human expressed pr
5	2338	99.8	435	20	Human protease HUP
6	2338	99.8	435	24	Human expressed pr
7	2337	99.8	437	23	CJA8 preferred seq
8	2337	99.8	437	24	Human expressed pr
9	2337	99.8	437	24	Human expressed pr

10	2337	99.8	437	24	ABU04935	Human expressed pr
11	2337	99.8	437	24	ABU04936	Human expressed pr
12	2329	99.4	492	22	AAU72559	Human seripancrin
13	2329	99.4	492	24	ABU04913	Human expressed pr
14	2329	99.4	492	24	ABU04932	Human expressed pr
15	2319	99.0	437	22	AAE06931	Human membrane-ty
16	2319	99.0	437	24	ABU04917	Human expressed pr
17	2319	99.0	437	24	ABU04922	Human expressed pr
18	2297.5	98.1	432	21	AAU99417	Human PRO1570 (UNQ
19	2297.5	98.1	432	22	AAU29188	Human PRO1570. Ho
20	2297.5	98.1	432	22	AAU87581	Protein of the inv
21	2297.5	98.1	432	22	AAU66166	Human secreted/tra
22	2297.5	98.1	432	23	ASG95906	Tumour-associated
23	2297.5	98.1	432	23	AAU76535	Human PRO1570 prot
24	2297.5	98.1	432	24	ABU71276	Human secreted pol
25	2297.5	98.1	432	24	ABU71561	Novel human secret
26	2297.5	98.1	432	24	ABU72007	Human PRO polypept
27	2297.5	98.1	432	24	ABU72164	Human secreted/tra
28	2297.5	98.1	432	24	ABU65733	Human secreted/tra
29	2297.5	98.1	432	24	ABU66066	Novel human secret
30	2297.5	98.1	432	24	ABU67570	Human secreted/tra
31	2297.5	98.1	432	24	ABU65428	Human PRO polypept
32	2297.5	98.1	432	24	ABU58564	Human PRO polypept
33	2297.5	98.1	432	24	ABU56100	Human secreted/tra
34	2297.5	98.1	432	24	ABU57095	Human PRO polypept
35	2297.5	98.1	432	24	ABU04921	Human expressed pr
36	2297.5	98.1	432	24	ABU04929	Human expressed pr
37	2297.5	98.1	432	24	ABU04934	Human expressed pr
38	2297.5	98.1	432	24	ABU10674	Human secreted/tra
39	2287	97.7	428	23	ABG96430	Human ovarian can
40	2251	96.1	521	22	AAU16975	Human novel secret
41	2251	96.1	521	24	ABU04925	Human expressed pr
42	2242	95.7	420	24	ABU56625	Lung cancer-associ
43	2198	93.9	423	22	AAE06944	Human transmembran
44	2198	93.9	423	24	ABU04918	Human expressed pr
45	2198	93.9	423	24	ABU04919	Human expressed pr

ALIGNMENTS

RESULT 1

AAU72558

ID AAU72558 standard; Protein; 435 AA.

XX

AC AAU72558;

XX

02-MAY-2001 (first entry)

XX

DT Human seripancrin protein.

DE

XX

Human; seripancrin; serine protease; chromosome 11q22-q23; therapy; arthritis; chronic obstructive pulmonary disorder; COPD; cancer; osteoporosis; aberrant wound healing; angiogenesis; diabetes; inflammatory disorder; stroke; cardiovascular disease; gene therapy; vaccine; cytostatic; cerebroprotective; vulnary; osteopathic.

OS Homo sapiens.

XX

PN WO200104141-A2.

XX

PD 18-JAN-2001.

XX

PF 04-JUL-2000; 2000WO-EF06211.

XX

PR 12-JUL-1999; 99EP-0113428.

XX

PA (MERE) MERCK PATENT GMBH.

XX

PI Suendermann B, Hofmann U, Matzku S, Wilbert O;

XX

DR WPI: 2001-147177/15.

DR N-PSDB; AAD02556.

```
XX PT New extracellular serine protease Seripancrin, useful for treating
PT cancer, osteoporosis, arthritis, chronic obstructive pulmonary disease,
PT diabetes, inflammatory disorders, stroke, angiogenesis and aberrant
PT wound healing
XX
PS Claim 2; Page 39-40; 45pp; English.
XX
XX The present invention relates to seripancrin polynucleotides, and
XX polypeptides encoded by them. Seripancrin are members of serine
XX protease family. This protein contains a transmembrane domain,
XX a low density lipoprotein (LDL) domain, protease domain and a
XX scavenger receptor cysteine-rich (SCR) domain. The LDL and SCR
XX domains help to define the specificity of seripancrin's intra and
XX intermolecular interactions. The polynucleotides and polypeptides
XX of the invention are useful for treating and diagnosing diseases
XX such as arthritis, chronic obstructive pulmonary disorder (COPD),
XX cancer, osteoporosis, aberrant wound healing, angiogenesis,
XX inflammatory disorders, diabetes, stroke and cardiovascular diseases.
XX Seripancrin genes are useful in chromosome localisation studies,
XX as tools for tissue expression studies and also in gene therapy.
XX The polypeptides of the invention are used for identifying agonists
XX and antagonists useful for treating conditions associated with
XX seripancrin imbalance. These polypeptides are also useful as vaccines.
XX The present sequence is seripancrin protein. The
XX seripancrin gene is located on human chromosome 11q22-q23.
XX
SQ Sequence 435 AA;
Query Match 100.0%; Score 2342; DB 22; Length 435;
Best Local Similarity 100.0%; Pred. No. 3.3e-194;
Matches 435; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MDPDSQPLNSLDVKPLRKPRIPMETPRKVGIPITIIALLSLASIIIVWLKVKLYKF 60
Db 1 MDPDSQPLNSLDVKPLRKPRIPMETPRKVGIPITIIALLSLASIIIVWLKVKLYKF 60
QY 61 LCGOPLHFIPRKQICGELDCPLGEDBEHCVKFPPEGPAVAVRLSKDRSTLQVLD SATGN 120
Db 61 LCGOPLHFIPRKQICGELDCPLGEDBEHCVKFPPEGPAVAVRLSKDRSTLQVLD SATGN 120
QY 121 WFSACFDNFTTEALAEATACRMGYSSTKPTFAVEIGPDODLDVVVEITENSQELMRNSSGP 180
Db 121 WFSACFDNFTTEALAEATACRMGYSSTKPTFAVEIGPDODLDVVVEITENSQELMRNSSGP 180
QY 181 CLSGSLVSLHCLACGSKLTPRVVVGGEASVDSWPQVSIQYDKQHVCGGSIIDPHWLT 240
Db 181 CLSGSLVSLHCLACGSKLTPRVVVGGEASVDSWPQVSIQYDKQHVCGGSIIDPHWLT 240
QY 241 AAHCPRKHTDVFNNKVRAGSKDLGSPSPSLAVAKIIIEFNPMYPKNDIATMLKQPLTF 300
Db 241 AAHCPRKHTDVFNNKVRAGSKDLGSPSPSLAVAKIIIEFNPMYPKNDIATMLKQPLTF 300
QY 301 SGTVRPICLPFPDEELTPATPLWIIWGFTKONGKMSDILLQASVQVIDSTECNADDAY 360
Db 301 SGTVRPICLPFPDEELTPATPLWIIWGFTKONGKMSDILLQASVQVIDSTECNADDAY 360
QY 361 QGEVTEKMKCAGIPEGGVDTCCGSDGSPLMYQSDQHWVGVISWGVGCGGSPTPGYTKV 420
Db 361 QGEVTEKMKCAGIPEGGVDTCCGSDGSPLMYQSDQHWVGVISWGVGCGGSPTPGYTKV 420
QY 421 SAYLNWIYNVWKAEL 435
Db 421 SAYLNWIYNVWKAEL 435
RESULT 2
AAG78577
ID AAG78577 standard; Protein; 435 AA.
XX
XX AAG78577;
XX
XX 07-MAY-2002 (first entry)
DT
```

```
XX DE Human serine protease D-G amino acid sequence.
XX
XX Serine protease; D-G; human; zymogen; enzyme; cytostatic;
XX antiinflammatory; dermatological; anticoagulation; cancer;
XX skin disorder; neuropathic pain; inflammatory disorder;
XX coagulation diathesis; thrombosis; laundry detergent; skin care;
XX gene therapy.
XX
XX Homo sapiens.
OS
FH Key Location/Qualifiers
FH Region 31..52
FT /note= "hydrophobic transmembrane domain"
FT Active-site 202..203
FT /note= "residues spanning the zymogen cleavage site"
FT Active-site 243
FT /note= "catalytic triad residue"
FT Active-site 339
FT /note= "catalytic triad residue"
FT Misc-difference 361
FT /note= "Encoded by CTG"
FT Active-site 385
FT /note= "catalytic triad residue"
XX WO200202011-A1.
XX
XX 10-JAN-2002.
XX
XX 08-JUN-2001; 2001WO-US18568.
XX
XX 30-JUN-2000; 2000US-0607745.
XX
XX (ORTH ) ORTHO-MCNEIL PHARM INC.
XX
XX Darrow AL, Qi J, Andrade-Gordon P;
XX
XX WPI; 2002-106601/14.
XX N-PSDB; AAI64284.
XX
XX Nucleic acid encoding a serine protease called D-G protein which is
XX useful for identifying modulators that are useful for treating a
XX condition which is mediated by protease D-G, e.g. cancer, skin
XX disorders, or neuropathic pain
XX
XX Claim 13; Fig 1B; 81pp; English.
XX
XX The invention relates to an isolated and purified nucleic acid that
XX encodes a serine protease called D-G protein. The activity of the protein
XX of the invention may be described as cytostatic, antiinflammatory,
XX dermatological and anticoagulation. The serine protease of the invention
XX is a member of the trypsin/chymotrypsin-like (SI) serine protease family,
XX which play an important role in processes such as digestion and
XX regulatory amplification cascades through the proteolytic activation of
XX inactive zymogen precursors. Protease D-G modulating compounds are useful
XX for treating a condition which is mediated by protease D-G, e.g. cancer,
XX skin disorders, neuropathic pain, inflammatory disorders, or coagulation
XX diathesis/thrombosis. The polynucleotide encoding the protease is useful
XX for identifying, detecting or isolating mutant forms of DNA molecules
XX encoding the protease. The protease is useful for identifying modulators
XX of the functional protease. The D-G protein can be used for formulation
XX of compositions for laundry detergents and skin care products. Protease
XX D-G gene therapy may be used to introduce protease D-G into the cells of
XX target organisms. As the D-G protein is derived from a human, it is less
XX likely to produce an allergic reaction in sensitive individuals when used
XX in formulations for laundry detergents and skin care products. The
XX current sequence represents the human serine protease D-G amino acid
XX sequence.
SQ Sequence 435 AA;
Query Match 100.0%; Score 2342; DB 23; Length 435;
Best Local Similarity 100.0%; Pred. No. 3.3e-194;
```

Matches 435; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDPDSQPLNSLDVPLKPRIPMETFRKVGIPITIALSLASIIIVVLIKVLIDKYTF 60
 DB |||||
 QY 1 MDPDSQPLNSLDVPLKPRIPMETFRKVGIPITIALSLASIIIVVLIKVLIDKYTF 60
 DB |||||

QY 61 LCGQPLHFIIPRKQDCGELDCPLGDEEHCVKSPFPGPAVAVRLSKDRSTLOVLD SATGN 120
 DB |||||

QY 61 LCGQPLHFIIPRKQDCGELDCPLGDEEHCVKSPFPGPAVAVRLSKDRSTLOVLD SATGN 120
 DB |||||

QY 121 WFSACFDNTEALAEATACRQMGYSKPTFRVIGPDQDLDVVEITENSQELMRNSSGP 180
 DB |||||

QY 121 WFSACFDNTEALAEATACRQMGYSKPTFRVIGPDQDLDVVEITENSQELMRNSSGP 180
 DB |||||

QY 181 CLSGSLVSLHCLACGKSLKTRPVVGEEASVDSWPQVSIQYDKQHVCGSILDPHWLT 240
 DB |||||

QY 181 CLSGSLVSLHCLACGKSLKTRPVVGEEASVDSWPQVSIQYDKQHVCGSILDPHWLT 240
 DB |||||

QY 241 AAHCFRKHTDVFNMKVRAGSKLGSPFSLAVAKIIIEFNPMYPKNDIALMKLOPLTF 300
 DB |||||

QY 241 AAHCFRKHTDVFNMKVRAGSKLGSPFSLAVAKIIIEFNPMYPKNDIALMKLOPLTF 300
 DB |||||

QY 301 SGTVRPICLPFFDEELTPATPLIIGWFTKQNGKMSDILLQASVQVIDSTRCNADDDAY 360
 DB |||||

QY 301 SGTVRPICLPFFDEELTPATPLIIGWFTKQNGKMSDILLQASVQVIDSTRCNADDDAY 360
 DB |||||

QY 361 QGEVTEKMKCAGIPGEGVDTCCQDSGGPLMYQSDQHVHVGIVSWGVCSTPGVYTKV 420
 DB |||||

QY 361 QGEVTEKMKCAGIPGEGVDTCCQDSGGPLMYQSDQHVHVGIVSWGVCSTPGVYTKV 420
 DB |||||

QY 421 SAYLNWIYNVWKAEL 435
 DB |||||

QY 421 SAYLNWIYNVWKAEL 435
 DB |||||

RESULT 3
 ABU04912
 ID ABU04912 standard; Protein; 435 AA.
 XX AC ABU04912;
 XX DT 29-JAN-2003 (first entry)
 XX XX

29-JAN-2003 (first entry)
 Human expressed protein tag (EPT) #1578.
 Translational profiling; expressed protein tag; EPT; kinase;
 phosphatase; protease; protease inhibitor; transporter;
 cytoskeletal protein; receptor; transcription factor; cancer; MHC;
 major histocompatibility complex; myeloma; colon cancer;
 gastric cancer; adenocarcinoma; sarcoma; melanoma; lymphoma;
 leukaemia.

Homo sapiens.
 OS XX
 NC XX
 WO200278524-A2.
 XX PN
 XX XX
 10-OCT-2002.
 XX PF
 XX PF
 28-MAR-2002; 2002WO-US09671.
 XX PF
 XX PF
 28-MAR-2001; 2001US-279495P.
 PR PR
 21-MAY-2001; 2001US-292544P.
 PR PR
 08-AUG-2001; 2001US-310801P.
 PR PR
 01-OCT-2001; 2001US-326370P.
 PR PR
 04-DEC-2001; 2001US-336780P.
 PR PR
 20-FEB-2002; 2002US-358985P.
 XX XX
 (ZYCO-) ZYCOS INC.
 PA PA
 XX XX
 XX XX
 Chiciz RM, Tomlinson AJ, Urban RG;
 PI PI
 XX XX
 WPI; 2003-040607/03.
 DR DR
 XX XX

PT New polypeptides (e.g. kinases, phosphatases, proteases, transporters, cytoskeletal proteins, receptors or transcription factors), useful for treating cancer, e.g. colon cancer, gastric cancer, sarcoma, lymphoma or leukemia

PS Example 2; SEQ ID No 1578; 134pp; English.

XX

CC The invention describes a purified polypeptide, which comprises a fragment of a kinase, phosphatase, protease, protease inhibitor, transporter, cytoskeletal protein, receptor or transcription factor. The polypeptide is useful as an immunogenic composition for eliciting in a mammal an immunogenic response directed against any of the purified polypeptide. The purified polypeptide, or the antibody that binds to this polypeptide, is useful for treating cancer. The polypeptide is also useful for identifying compounds that binds to a naturally processed class I or class II MHC-binding polypeptide. The polypeptides and polynucleotides are particularly useful for treating or preventing myeloma, colon cancer, gastric cancer, adenocarcinoma, sarcoma, melanoma, lymphoma or leukemia. These are also useful for screening agents for treating the above mentioned diseases. This sequence represents an expressed protein tag (EPT) isolated from human tissue for translational profiling.

CC Note: This sequence does not appear in the printed specification but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

XX

SQ Sequence 435 AA;

Query Match 100.0%; Score 2342; DB 24; Length 435;
 Best Local Similarity 100.0%; Pred. No. 3.3e-194;
 Matches 435; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDPDSQPLNSLDVPLKPRIPMETFRKVGIPITIALSLASIIIVVLIKVLIDKYTF 60
 DB 1 MDPDSQPLNSLDVPLKPRIPMETFRKVGIPITIALSLASIIIVVLIKVLIDKYTF 60

QY 61 LCGQPLHFIIPRKQDCGELDCPLGDEEHCVKSPFPGPAVAVRLSKDRSTLOVLD SATGN 120
 DB 61 LCGQPLHFIIPRKQDCGELDCPLGDEEHCVKSPFPGPAVAVRLSKDRSTLOVLD SATGN 120

QY 121 WFSACFDNTEALAEATACRQMGYSKPTFRVIGPDQDLDVVEITENSQELMRNSSGP 180
 DB 121 WFSACFDNTEALAEATACRQMGYSKPTFRVIGPDQDLDVVEITENSQELMRNSSGP 180

QY 181 CLSGSLVSLHCLACGKSLKTRPVVGEEASVDSWPQVSIQYDKQHVCGSILDPHWLT 240
 DB 181 CLSGSLVSLHCLACGKSLKTRPVVGEEASVDSWPQVSIQYDKQHVCGSILDPHWLT 240

QY 241 AAHCFRKHTDVFNMKVRAGSKLGSPFSLAVAKIIIEFNPMYPKNDIALMKLOPLTF 300
 DB 241 AAHCFRKHTDVFNMKVRAGSKLGSPFSLAVAKIIIEFNPMYPKNDIALMKLOPLTF 300

QY 301 SGTVRPICLPFFDEELTPATPLIIGWFTKQNGKMSDILLQASVQVIDSTRCNADDDAY 360
 DB 301 SGTVRPICLPFFDEELTPATPLIIGWFTKQNGKMSDILLQASVQVIDSTRCNADDDAY 360

QY 361 QGEVTEKMKCAGIPGEGVDTCCQDSGGPLMYQSDQHVHVGIVSWGVCSTPGVYTKV 420
 DB 361 QGEVTEKMKCAGIPGEGVDTCCQDSGGPLMYQSDQHVHVGIVSWGVCSTPGVYTKV 420

QY 421 SAYLNWIYNVWKAEL 435
 DB 421 SAYLNWIYNVWKAEL 435

RESULT 4
 ABU04931
 ID ABU04931 standard; Protein; 435 AA.
 XX AC ABU04931;
 XX DT 29-JAN-2003 (first entry)
 XX XX

Human expressed protein tag (EPT) #1597.

Translational profiling; expressed protein tag; EPT; kinase; phosphatase; protease; protease inhibitor; transporter; cytoskeletal protein; receptor; transcription factor; cancer; MHC; major histocompatibility complex; myeloma; colon cancer; gastric cancer; adenocarcinoma; sarcoma; melanoma; lymphoma; leukaemia.

Homo sapiens.

W0200278524-A2.

10-OCT-2002.

28-MAR-2002; 2002WO-US09671.

28-MAR-2001; 2001US-279495P.

21-MAY-2001; 2001US-292544P.

08-AUG-2001; 2001US-310801P.

01-OCT-2001; 2001US-326370P.

04-DEC-2001; 2001US-336780P.

20-FEB-2002; 2002US-358985P.

(ZYCO-) ZYCOS INC.

Chicz RM, Tomlinson AJ, Urban RG;

WPI; 2003-040607/03.

New polypeptides (e.g. kinases, phosphatases, proteases, transporters, cytoskeletal proteins, receptors or transcription factors), useful for treating cancer, e.g. colon cancer, gastric cancer, sarcoma, lymphoma or leukemia

Example 2; SEQ ID No 1597; 134pp; English.

The invention describes a purified polypeptide, which comprises a fragment of a kinase, phosphatase, protease, protease inhibitor, transporter, cytoskeletal protein, receptor or transcription factor. The polypeptide is useful as an immunogenic composition for eliciting in a mammal an immunogenic response directed against any of the purified polypeptide. The purified polypeptide, or the antibody that binds to this polypeptide, is useful for treating cancer. The polypeptide is also useful for identifying compounds that binds to a naturally processed class I or class II MHC-binding polypeptide. The polypeptides and polynucleotides are particularly useful for treating or preventing myeloma, colon cancer, gastric cancer, adenocarcinoma, sarcoma, melanoma, lymphoma or leukaemia. These are also useful for screening agents for treating the above mentioned diseases. This sequence represents an expressed protein tag (EPT) isolated from human tissue for translational profiling.

Note: This sequence does not appear in the printed specification but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

Sequence 435 AA;

Query Match 100.0%; Score 2342; DB 24; Length 435;
Best Local Similarity 100.0%; Pred. No. 3.3e-194;
Matches 435; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MDPDSQPLNSLDVPLKPRIPMETFRKVGIPPIIIALLSLASIIIVVLIKVLDKYIF 60
1 MDPDSQPLNSLDVPLKPRIPMETFRKVGIPPIIIALLSLASIIIVVLIKVLDKYIF 60
61 LCGQPLHFIPLKQDCELDCEPGEDEECVKSFPPEGPAVAVRLSKDRSTLQVLDSATGN 120
61 LCGQPLHFIPLKQDCELDCEPGEDEECVKSFPPEGPAVAVRLSKDRSTLQVLDSATGN 120
121 WFSACPDNTEALAEATACQMGYSSKPTFRAVEIGPDQDLDDVVEITENSQELMRNSSGP 180
121 WFSACPDNTEALAEATACQMGYSSKPTFRAVEIGPDQDLDDVVEITENSQELMRNSSGP 180

QY	181	CLSGSIVSLHCLACGSKLTPRVVGGEEASVDSWPQVSIQYDKQHVCGSILDPHWLT	240
Db	181	CLSGSIVSLHCLACGSKLTPRVVGGEEASVDSWPQVSIQYDKQHVCGSILDPHWLT	240
QY	241	AAHCFRKHTDVFNWKRAGSDKLGSPFSLAVAKIIIEFNPMYPKNDIALMKLQPLTF	300
Db	241	AAHCFRKHTDVFNWKRAGSDKLGSPFSLAVAKIIIEFNPMYPKNDIALMKLQPLTF	300
QY	301	SGTVREICLPFFDEELTPATPLWIIWGFTKONGGKMSDILLQASVQVIDSTRCNADDA	360
Db	301	SGTVREICLPFFDEELTPATPLWIIWGFTKONGGKMSDILLQASVQVIDSTRCNADDA	360
QY	361	QGEVTEKMKCAGIPGEGVDTCCGDSGGPLMYQSDQHHVVGIVSWGYCGGSPGVTYTKV	420
Db	361	QGEVTEKMKCAGIPGEGVDTCCGDSGGPLMYQSDQHHVVGIVSWGYCGGSPGVTYTKV	420
QY	421	SAYLNWIYNVWKAEI 435	
Db	421	SAYLNWIYNVWKAEI 435	
RESULT 5			
ID	AA106437	standard; Protein; 435 AA.	
XX	AC	AA106437;	
XX	DT	27-SEP-1999 (first entry)	
XX	DE	Human protease HUPM-6.	
XX	KW	Serine protease; human; HUPM-6; cell proliferation; cancer;	
XX	KW	immune disorder; inflammation; therapy.	
XX	OS	Homo sapiens.	
XX	PH	Key	Location/Qualifiers
XX	FT	Modified-site	26
XX	FT	Modified-site	59
XX	FT	Modified-site	93
XX	FT	Modified-site	128
XX	FT	Modified-site	144
XX	FT	Modified-site	148
XX	FT	Modified-site	176
XX	FT	Modified-site	197
XX	FT	Modified-site	200
XX	FT	Modified-site	231
XX	FT	Active-site	243
XX	FT	Modified-site	249
XX	FT	Modified-site	260
XX	FT	Modified-site	303
XX	FT	Modified-site	351
XX	FT	Modified-site	360
XX	FT	Modified-site	365
XX	FT	Modified-site	

FT Active-site 385
ET /note= "Characteristic of serine protease"
XT
FN WO9936550-A2.
XX
XX 22-JUL-1999.
XX
XX 12-JAN-1999; 99WO-US00655.
XX
XX 16-JAN-1999; 98US-0008271.
XX
XX (INCY-) INCYTE PHARM INC.
XX
XX Bandman O, Corley NC, Guegler KJ, Hillman JL, Shah P;
XX Tang YT, Yue H;
XX WPI; 1999-430616/36.
XX N-PSDB; AAX87154.
XX
XX Novel human protease molecules useful in the treatment of
XX developmental disorders and/or cancers
XX
XX Claim 1; Page 74-75; 90pp; English.
XX
XX The present sequence represents novel human protease HUPM-6, as
XX deduced from the consensus sequence (see AAX87154) of overlapping
XX cDNA clones obtained from various libraries. Northern analysis
XX shows expression of HUPM-6 in gastrointestinal, and male and
XX female reproductive cDNA libraries. Approximately 65% of these
XX libraries are associated with neoplastic disorders, and 22% with
XX the immune response. The invention provides 12 new human
XX proteases, i.e. HUPM-1 to -12 (see AAX06432-43), and the
XX polynucleotides encoding them (see AAX87149-60). Also provided are
XX vectors, host cells and methods for producing HUPM polypeptides,
XX as well as agonists and antagonists of HUPM. Methods for treating
XX or preventing cell proliferative disorders and immune disorders
XX using HUPM or HUPM antagonists are claimed.
XX
XX Sequence 435 AA;
XX
XX Query Match 99.8%; Score 2338; DB 20; Length 435;
XX Best Local Similarity 99.8%; Pred. No. 7.3e-194;
XX Matches 434; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
XX
XX 1 MDPSDQPLNSLDVPLRPRIPMETFRKVGPIIIALLSLASIIIVVLKVIDKYIF 60
XX
XX 1 MDPSDQPLNSLDVPLRPRIPMETFRKVGPIIIALLSLASIIIVVLKVIDKYIF 60
XX
XX 61 LCGQPLHFIIPRQKLDGELDCPLGDEDEHCVKSPFEGPAVAVRLSKDRSTLQVLD SATGN 120
XX
XX 61 LCGQPLHFIIPRQKLDGELDCPLGDEDEHCVKSPFEGPAVAVRLSKDRSTLQVLD SATGN 120
XX
XX 121 WFSACFDNFTALAEATACRMGYSKPTFRAVEIGPDQDLVVEITENSQELMRNSGP 180
XX
XX 121 WFSACFDNFTALAEATACRMGYSKPTFRAVEIGPDQDLVVEITENSQELMRNSGP 180
XX
XX 181 CLSGSLVSHLCLACGSKLTPRWGGEERASVSWPQVSIQDKHQHVCVGSILDPHWILT 240
XX
XX 181 CLSGSLVSHLCLACGSKLTPRWGGEERASVSWPQVSIQDKHQHVCVGSILDPHWILT 240
XX
XX 241 AAHCFRKHDTFVNKVRAGSDKLGSPFSLAVAKIIIEFNPMYPKDNIALMKLOFPITF 300
XX
XX 241 AAHCFRKHDTFVNKVRAGSDKLGSPFSLAVAKIIIEFNPMYPKDNIALMKLOFPITF 300
XX
XX 301 SGTVPICLPFFDELTPTPLWIIIGWGTCKQGGKMSDILLQASVQVIDSTRCNADDAY 360
XX
XX 301 SGTVPICLPFFDELTPTPLWIIIGWGTCKQGGKMSDILLQASVQVIDSTRCNADDAY 360
XX
XX 361 QGEVTEKMKACIPGEGVDTCGDSGGPLMYQSDQHVHVGIVSWGCGGPGTPGYTKV 420
XX
XX 361 QGEVTEKMKACIPGEGVDTCGDSGGPLMYQSDQHVHVGIVSWGCGGPGTPGYTKV 420
XX
XX 421 SAYLWNIYNVWKAEL 435

Db 421 SAYLWNIYNVWKAEL 435
|||||
RESULT 6
ABU04930
ID ABU04930 standard; Protein; 435 AA.
XX
XX AC ABU04930;
XX
XX 29-JAN-2003 (first entry)
XX
XX Human expressed protein tag (EPT) #1596.
XX
XX Translational profiling; expressed protein tag; EPT; kinase;
XX phosphatase; protease; protease inhibitor; transporter;
XX cytoskeletal protein; receptor; transcription factor; cancer; MHC;
XX major histocompatibility complex; myeloma; colon cancer;
XX gastric cancer; adenocarcinoma; sarcoma; melanoma; lymphoma;
XX leukaemia.
XX
XX Homo sapiens.
XX
XX WO200278524-A2.
XX
XX 10-OCT-2002.
XX
XX 28-MAR-2002; 2002WO-US09671.
XX
XX 28-MAR-2001; 2001US-279495P.
XX
XX 21-MAY-2001; 2001US-292544P.
XX
XX 08-AUG-2001; 2001US-310801P.
XX
XX 01-OCT-2001; 2001US-326370P.
XX
XX 04-DEC-2001; 2001US-336780P.
XX
XX 20-FEB-2002; 2002US-358985P.
XX
XX (ZYCO-) ZYCOS INC.
XX
XX Chicx RM, Tomlinson AJ, Urban RG;
XX
XX WPI; 2003-040607/03.
XX
XX New polypeptides (e.g. kinases, phosphatases, proteases, transporters,
XX cytoskeletal proteins, receptors or transcription factors), useful for
XX treating cancer, e.g. colon cancer, gastric cancer, sarcoma, lymphoma
XX or leukemia
XX
XX Example 2; SEQ ID No 1596; 134pp; English.
XX
XX The invention describes a purified polypeptide, which comprises a
XX fragment of a kinase, phosphatase, protease, protease inhibitor,
XX transporter, cytoskeletal protein, receptor or transcription factor.
XX The polypeptide is useful as an immunogenic composition for eliciting
XX in a mammal an immunogenic response directed against any of the purified
XX polypeptide. The purified polypeptide, or the antibody that binds to
XX this polypeptide, is useful for treating cancer. The polypeptide is
XX also useful for identifying compounds that binds to a naturally
XX processed class I or class II MHC-binding polypeptide. The polypeptides
XX and polynucleotides are particularly useful for treating or preventing
XX myeloma, colon cancer, gastric cancer, adenocarcinoma, sarcoma, melanoma,
XX lymphoma or leukaemia. These are also useful for screening agents for
XX treating the above mentioned diseases. This sequence represents an
XX expressed protein tag (EPT) isolated from human tissue for translational
XX profiling.
XX Note: This sequence does not appear in the printed specification but was
XX obtained in electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 435 AA;
XX
XX Query Match 99.8%; Score 2338; DB 24; Length 435;
XX Best Local Similarity 99.8%; Pred. No. 7.3e-194;
XX Matches 434; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDPDSQPLNSLDVKPLRKPRIPMETFRKVGIPITIIALLSLASIIIVVLIKVLDRKYF 60
 Db |||||
 QY 1 MDPDSQPLNSLDVKPLRKPRIPMETFRKVGIPITIIALLSLASIIIVVLIKVLDRKYF 60
 Db |||||
 QY 61 LCGQPLHFIPRKQCDGELDCPLGEDEEHCVKSPFPGPAVAVLSKDRSTLOVLDSTGN 120
 Db |||||
 QY 61 LCGQPLHFIPRKQCDGELDCPLGEDEEHCVKSPFPGPAVAVLSKDRSTLOVLDSTGN 120
 Db |||||
 QY 121 WFSACFDNFTALAEACRQMGYSKPTFRAVEIGPDQDLVDVEITENSQELMRNSSGP 180
 Db |||||
 QY 121 WFSACFDNFTALAEACRQMGYSKPTFRAVEIGPDQDLVDVEITENSQELMRNSSGP 180
 Db |||||
 QY 181 CLSGSLVSLHCLACGSLKTPRVVGGSEASVDSWPQVSIQYDKQHVCGGSIIDPHWLT 240
 Db |||||
 QY 181 CLSGSLVSLHCLACGSLKTPRVVGGSEASVDSWPQVSIQYDKQHVCGGSIIDPHWLT 240
 Db |||||
 QY 241 AAHCFRKHDTDFNWKVRAGSDKLSRFPFLAVAKIIIEFNPMYPKNDIALMKLOPPLTF 300
 Db |||||
 QY 241 AAHCFRKHDTDFNWKVRAGSDKLSRFPFLAVAKIIIEFNPMYPKNDIALMKLOPPLTF 300
 Db |||||
 QY 301 SGTVRPICLPFFDEELTPATPLMIIGWFTKONGKMSDILLOASQVVIDSTRCNADDA 360
 Db |||||
 QY 301 SGTVRPICLPFFDEELTPATPLMIIGWFTKONGKMSDILLOASQVVIDSTRCNADDA 360
 Db |||||
 QY 361 QGVTETKMCAGIPEGGVDTCCGDSGGPLMYQSDQHVHVGIVSWGVCSTPGVYTKV 420
 Db |||||
 QY 361 QGVTETKMCAGIPEGGVDTCCGDSGGPLMYQSDQHVHVGIVSWGVCSTPGVYTKV 420
 Db |||||
 QY 421 SAYLNWIYNVWKAEL 435
 Db |||||
 QY 421 SAYLNWIYNVWKAEL 435
 Db |||||

RESULT 7
 AAG79359
 ID AAG79359 standard; Protein; 437 AA.
 XX
 AAG79359;
 XX

21-AUG-2002 (first entry)

CJA8 preferred sequence.

Colorectal cancer; CGA7; CJA8; modulating protein; screening; drug candidate; vaccine.

Homo sapiens.

Key Location/Qualifiers
 Domain 36..63
 /note= "Transmembrane domain"

US2002042067-A1.

11-APR-2002.

08-MAY-2001; 2001US-0851588.

17-AUG-2000; 2000US-0642252.

06-SEP-2000; 2000US-0656002.

(MACK/) MACK D.

(GISH/) GISH K C.

(WILS/) WILSON K E.

Mack D, Gish KC, Wilson KE;

WPI; 2002-453647/48.

N-PSDB; AAI72976.

Screening drug candidates for treating colorectal cancer, comprises
 determining the effect of the candidate on the expression profile gene

PT of CGA7 or CJA8 -
 XX Disclosure; Fig 8; 40pp; English.
 PS
 XX
 CC The sequences given in AAG79356-59 show the colorectal cancer
 CC proteins, CGA7 and CJA8. The CGA7 and CJA8 proteins are colorectal
 CC cancer modulating proteins and have been mapped to chromosomes 2 (CGA7)
 CC and 11 (CJA8). These sequences may be used in the method of the
 CC invention for screening drug candidates. The method comprises adding
 CC a drug candidate to a cell that expresses an expression profile gene
 CC encoding CGA7, CJA8 or fragments and determining the effect of the
 CC drug candidate on the expression of the expression profile gene. The new
 CC methods are used to screen bioactive agents for the ability to bind to
 CC or modulate the activity of CGA7 or CJA8 and evaluate the effect of a
 CC candidate colorectal cancer drug. An antibody to CGA7 or CJA8 can inhibit
 CC the activity of CGA7 or CJA8, respectively, and is used to screen for
 CC an agent that can interfere with the binding of CGA7 or CJA8 to the
 CC antibody. The antibody can be used to treat colorectal cancer. The
 CC antibody or a fragment of it is used to localize a therapeutic group to
 CC a colorectal cancer tissue, where the therapeutic group is a cytotoxic
 CC agent or a radioisotope. Antisense molecules are used to inhibit
 CC colorectal cancer in a cell. Nucleic acid segments encoding CGA7 or
 CC CJA8 are used in a biochip. CGA7, CJA8 or a nucleic acid encoding it
 CC are used to elicit an immune response. CGA7 or CJA8 is used to
 CC determine the prognosis of an individual with colorectal cancer.
 CC Nucleic acid encoding CGA7 or CJA8 can be used in vaccines.
 XX
 SQ Sequence 437 AA;

Query Match 99.8%; Score 2337; DB 23; Length 437;
 Best Local Similarity 100.0%; Pred. No. 9e-194;
 Matches 434; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 2 DPDSQPLNSLDVKPLRKPRIPMETFRKVGIPITIIALLSLASIIIVVLIKVLDRKYF 61
 Db 4 DPDSQPLNSLDVKPLRKPRIPMETFRKVGIPITIIALLSLASIIIVVLIKVLDRKYF 63
 QY 62 CQGPLHFTPRKQCDGELDCPLGEDEEHCVKSPFPGPAVAVLSKDRSTLOVLDSTGN 121
 Db 64 CQGPLHFTPRKQCDGELDCPLGEDEEHCVKSPFPGPAVAVLSKDRSTLOVLDSTGN 123
 QY 122 FSACFDNFTALAEACRQMGYSKPTFRAVEIGPDQDLVDVEITENSQELMRNSSGP 181
 Db 124 FSACFDNFTALAEACRQMGYSKPTFRAVEIGPDQDLVDVEITENSQELMRNSSGP 183
 QY 182 LSGSLVSLHCLACGSLKTPRVVGGSEASVDSWPQVSIQYDKQHVCGGSIIDPHWLT 241
 Db 184 LSGSLVSLHCLACGSLKTPRVVGGSEASVDSWPQVSIQYDKQHVCGGSIIDPHWLT 243
 QY 242 AHCFRKHDTDFNWKVRAGSDKLSRFPFLAVAKIIIEFNPMYPKNDIALMKLOPPLTF 301
 Db 244 AHCFRKHDTDFNWKVRAGSDKLSRFPFLAVAKIIIEFNPMYPKNDIALMKLOPPLTF 303
 QY 302 GTVRPICLPFFDEELTPATPLMIIGWFTKONGKMSDILLOASQVVIDSTRCNADDA 361
 Db 304 GTVRPICLPFFDEELTPATPLMIIGWFTKONGKMSDILLOASQVVIDSTRCNADDA 363
 QY 362 QGVTETKMCAGIPEGGVDTCCGDSGGPLMYQSDQHVHVGIVSWGVCSTPGVYTKV 421
 Db 364 QGVTETKMCAGIPEGGVDTCCGDSGGPLMYQSDQHVHVGIVSWGVCSTPGVYTKV 423
 QY 422 AYLNWIYNVWKAEL 435
 Db 424 AYLNWIYNVWKAEL 437

RESULT 8

ABU04915

ID ABU04915 standard; Protein; 437 AA.

XX

AC ABU04915;

XX

DT 29-JAN-2003 (first entry)

XX DE Human expressed protein tag (EPT) #1581.

XX DE Translational profiling; expressed protein tag; EPT; kinase;

XX KW phosphatase; protease; protease inhibitor; transporter;

XX KW cytoskeletal protein; receptor; transcription factor; cancer; MHC;

XX KW major histocompatibility complex; myeloma; colon cancer;

XX KW gastric cancer; adenocarcinoma; sarcoma; melanoma; lymphoma;

XX KW leukaemia.

OS Homo sapiens.

XX PN WO200278524-A2.

XX PN 10-OCT-2002.

XX PF 28-MAR-2002; 2002WO-US09671.

XX PF 28-MAR-2001; 2001US-279495P.

XX PF 21-MAY-2001; 2001US-292544P.

XX PF 08-AUG-2001; 2001US-310801P.

XX PF 01-OCT-2001; 2001US-326370P.

XX PF 04-DEC-2001; 2001US-336780P.

XX PF 20-FEB-2002; 2002US-358985P.

XX (ZYCO-) ZYCOS INC.

XX Chicz RM, Tomlinson AJ, Urban RG;

XX WPI; 2003-040607/03.

XX New polypeptides (e.g. kinases, phosphatases, proteases, transporters,

XX PT cytoskeletal proteins, receptors or transcription factors), useful for

XX PT treating cancer, e.g. colon cancer, gastric cancer, sarcoma, lymphoma

XX PT or leukemia -

XX Example 2; SEQ ID No 1581; 134pp; English.

XX The invention describes a purified polypeptide, which comprises a

XX fragment of a kinase, phosphatase, protease, protease inhibitor,

XX transporter, cytoskeletal protein, receptor or transcription factor.

XX The polypeptide is useful as an immunogenic composition for eliciting

XX in a mammal an immunogenic response directed against any of the purified

XX polypeptide. The purified polypeptide, or the antibody that binds to

XX this polypeptide, is useful for treating cancer. The polypeptide is

XX also useful for identifying compounds that binds to a naturally

XX processed class I or class II MHC-binding polypeptide. The polypeptides

XX and polynucleotides are particularly useful for treating or preventing

XX myeloma, colon cancer, gastric cancer, adenocarcinoma, sarcoma, melanoma,

XX lymphoma or leukaemia. These are also useful for screening agents for

XX treating the above mentioned diseases. This sequence represents an

XX expressed protein tag (EPT) isolated from human tissue for translational

XX profiling.

XX Note: This sequence does not appear in the printed specification but was

XX obtained in electronic format directly from WIPO at

XX ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 437 AA;

XX Query Match 99.8%; Score 2337; DB 24; Length 437;

XX Best Local Similarity 100.0%; Pred. No. 9e-194;

XX Matches 434; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 DPDSQPLNSLDVKKPKRIPMETFRKVGIPFIALLSLASIIIVVLKILDKYPL 61

DB 4 DPDSQPLNSLDVKKPKRIPMETFRKVGIPFIALLSLASIIIVVLKILDKYPL 63

QY 62 CQQLHFIIRKQLCDGLPCIGEDDEHCKVSPFEGPFAVRLSKDRSTLQVLDSATGNW 121

DB 64 CQQLHFIIRKQLCDGLPCIGEDDEHCKVSPFEGPFAVRLSKDRSTLQVLDSATGNW 123

QY 122 FSACFDNFTALAEATACRQMGYSKPTFRAVEIGPDQDLVDVEITENSQELRMNSSGPC 181

DB 124 FSACFDNFTALAEATACRQMGYSKPTFRAVEIGPDQDLVDVEITENSQELRMNSSGPC 183

QY 182 LSGSLVSLHCLACGKSLKTPRVVGGEEASVDSPQVSIQYDKQHVCGSILDPHVLTA 241

DB 184 LSGSLVSLHCLACGKSLKTPRVVGGEEASVDSPQVSIQYDKQHVCGSILDPHVLTA 243

QY 242 AHCFRKHGTDVFNWKVRAGSDKLGSPFSLAVAKIIIEFNPMYPKNDIALMKLQFPPLTFS 301

DB 244 AHCFRKHGTDVFNWKVRAGSDKLGSPFSLAVAKIIIEFNPMYPKNDIALMKLQFPPLTFS 303

QY 302 GTVRPICLPFFDEELTPATPLWIIIGWFTKQNGKMSDILLQASVQVIDSTRCNADAYQ 361

DB 304 GTVRPICLPFFDEELTPATPLWIIIGWFTKQNGKMSDILLQASVQVIDSTRCNADAYQ 363

QY 362 GEVTEKMCAGIPEGGVDTCQDSCGGLMYQSDQHVHVGIVSWGCGGPGSPFPGVTVTKVS 421

DB 364 GEVTEKMCAGIPEGGVDTCQDSCGGLMYQSDQHVHVGIVSWGCGGPGSPFPGVTVTKVS 423

QY 422 AYLNWYNNWKAEL 435

DB 424 AYLNWYNNWKAEL 437

RESULT 9

ABU04920

ID ABU04920 standard; Protein; 437 AA.

XX AC ABU04920;

XX DT 29-JAN-2003 (first entry)

XX DE Human expressed protein tag (EPT) #1586.

XX KW Translational profiling; expressed protein tag; EPT; kinase;

XX KW phosphatase; protease; protease inhibitor; transporter;

XX KW cytoskeletal protein; receptor; transcription factor; cancer; MHC;

XX KW major histocompatibility complex; myeloma; colon cancer;

XX KW gastric cancer; adenocarcinoma; sarcoma; melanoma; lymphoma;

XX KW leukaemia.

XX OS Homo sapiens.

XX PN WO200278524-A2.

XX PD 10-OCT-2002.

XX PF 28-MAR-2002; 2002WO-US09671.

XX PF 28-MAR-2001; 2001US-279495P.

XX PF 21-MAY-2001; 2001US-292544P.

XX PF 08-AUG-2001; 2001US-310801P.

XX PF 01-OCT-2001; 2001US-326370P.

XX PF 04-DEC-2001; 2001US-336780P.

XX PF 20-FEB-2002; 2002US-358985P.

XX (ZYCO-) ZYCOS INC.

XX Chicz RM, Tomlinson AJ, Urban RG;

XX WPI; 2003-040607/03.

XX New polypeptides (e.g. kinases, phosphatases, proteases, transporters,

XX PT cytoskeletal proteins, receptors or transcription factors), useful for

XX PT treating cancer, e.g. colon cancer, gastric cancer, sarcoma, lymphoma

XX PT or leukemia -

XX Example 2; SEQ ID No 1586; 134pp; English.

XX The invention describes a purified polypeptide, which comprises a

XX fragment of a kinase, phosphatase, protease, protease inhibitor,

XX transporter, cytoskeletal protein, receptor or transcription factor.

XX The polypeptide is useful as an immunogenic composition for eliciting

XX in a mammal an immunogenic response directed against any of the purified

XX polypeptide. The purified polypeptide, or the antibody that binds to

XX this polypeptide, is useful for treating cancer. The polypeptide is

XX also useful for identifying compounds that binds to a naturally

XX processed class I or class II MHC-binding polypeptide. The polypeptides

XX and polynucleotides are particularly useful for treating or preventing

XX myeloma, colon cancer, gastric cancer, adenocarcinoma, sarcoma, melanoma,

XX lymphoma or leukaemia. These are also useful for screening agents for

XX treating the above mentioned diseases. This sequence represents an

XX expressed protein tag (EPT) isolated from human tissue for translational

XX profiling.

XX Note: This sequence does not appear in the printed specification but was

XX obtained in electronic format directly from WIPO at

XX ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 437 AA;

XX Query Match 99.8%; Score 2337; DB 24; Length 437;

XX Best Local Similarity 100.0%; Pred. No. 9e-194;

XX Matches 434; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 DPDSQPLNSLDVKKPKRIPMETFRKVGIPFIALLSLASIIIVVLKILDKYPL 61

DB 4 DPDSQPLNSLDVKKPKRIPMETFRKVGIPFIALLSLASIIIVVLKILDKYPL 63

QY 62 CQQLHFIIRKQLCDGLPCIGEDDEHCKVSPFEGPFAVRLSKDRSTLQVLDSATGNW 121

DB 64 CQQLHFIIRKQLCDGLPCIGEDDEHCKVSPFEGPFAVRLSKDRSTLQVLDSATGNW 123

QY 122 FSACFDNFTALAEATACRQMGYSKPTFRAVEIGPDQDLVDVEITENSQELRMNSSGPC 181

CC polypeptide. The purified polypeptide, or the antibody that binds to
 CC this polypeptide, is useful for treating cancer. The polypeptide is
 CC also useful for identifying compounds that binds to a naturally
 CC processed class I or class II MHC-binding polypeptide. The polypeptides
 CC and polynucleotides are particularly useful for treating or preventing
 CC myeloma, colon cancer, gastric cancer, adenocarcinoma, sarcoma, melanoma,
 CC lymphoma or leukaemia. These are also useful for screening agents for
 CC treating the above mentioned diseases. This sequence represents an
 CC expressed protein tag (EPT) isolated from human tissue for translational
 CC profiling.
 CC Note: This sequence does not appear in the printed specification but was
 CC obtained in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.
 XX
 XX

XX Sequence 437 AA;
 Query Match 99.8%; Score 2337; DB 24; Length 437;
 Best Local Similarity 100.0%; Pred. No. 9e-194;
 Matches 434; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 2 DPDSQPLNSLDVVKPLRPMPETFRKVGIPPIIIALLSLASIIIVVLIKVLIDKYFL 61
 Db 4 DPDSQPLNSLDVVKPLRPMPETFRKVGIPPIIIALLSLASIIIVVLIKVLIDKYFL 63
 QY 62 CGOPLHPIPRKQLCDGELDCPLGEDEEHCVKSPPEGPAVAVRLSKDRSTLQVLD SATGNW 121
 Db 64 CGOPLHPIPRKQLCDGELDCPLGEDEEHCVKSPPEGPAVAVRLSKDRSTLQVLD SATGNW 123
 QY 122 FSACFDNFTTEALAEACRQMGYSKPTFRAVEIGPDQDLVDWEITENSQELMRNSSGPC 181
 Db 124 FSACFDNFTTEALAEACRQMGYSKPTFRAVEIGPDQDLVDWEITENSQELMRNSSGPC 183
 QY 182 LSGSLVSLHCLACGKSLKTRPVVGGEEASVDSWPQVSIQYDKQHVCGSILDPHWWLTA 241
 Db 184 LSGSLVSLHCLACGKSLKTRPVVGGEEASVDSWPQVSIQYDKQHVCGSILDPHWWLTA 243
 QY 242 AHCFRKHDTVFNKVRAGSKLGSFSLAVAKIIIEFNPMYPKNDIALMKLQPLTFS 301
 Db 244 AHCFRKHDTVFNKVRAGSKLGSFSLAVAKIIIEFNPMYPKNDIALMKLQPLTFS 303
 QY 302 GTVRPCLPFDFBELTPATPLWIIIGWFTKQNGKMSDILLOASVQVIDSTRCNADDAYQ 361
 Db 304 GTVRPCLPFDFBELTPATPLWIIIGWFTKQNGKMSDILLOASVQVIDSTRCNADDAYQ 363
 QY 362 GEVTEKMGAGIPEGGVDTCCGSGPLMYQSDQHWVGVISWYGGGPGSTPGVTYKVS 421
 Db 364 GEVTEKMGAGIPEGGVDTCCGSGPLMYQSDQHWVGVISWYGGGPGSTPGVTYKVS 423
 QY 422 AYLNWIYNNVKAEI 435
 Db 424 AYLNWIYNNVKAEI 437

RESULT 10
 ABU04935
 ID ABU04935 standard; Protein; 437 AA.
 XX AC ABU04935;
 XX DT 29-JAN-2003 (first entry)
 XX DE Human expressed protein tag (EPT) #1601.
 XX KW Translational profiling; expressed protein tag; EPT; kinase;
 KW phosphatase; protease; protease inhibitor; transporter;
 KW cytoskeletal protein; receptor; transcription factor; cancer; MHC;
 KW major histocompatibility complex; myeloma; colon cancer;
 KW gastric cancer; adenocarcinoma; sarcoma; melanoma; lymphoma;
 KW leukaemia.
 XX OS Homo sapiens.
 XX PN WO200278524-A2.

XX 10-OCT-2002.
 XX 28-MAR-2002; 2002WO-US09671.
 XX 28-MAR-2001; 2001US-279495P.
 PR 21-MAY-2001; 2001US-292544P.
 PR 08-AUG-2001; 2001US-310801P.
 PR 01-OCT-2001; 2001US-326370P.
 PR 04-DEC-2001; 2001US-336780P.
 PR 20-FEB-2002; 2002US-358985P.
 XX (ZYCO-) ZYCOS INC.
 XX
 XX Chiciz RM, Tomlinson AJ, Urban RG;
 XX WPI; 2003-040607/03.
 DR
 XX New polypeptides (e.g. kinases, phosphatases, proteases, transporters,
 FT cytoskeletal proteins, receptors or transcription factors), useful for
 FT treating cancer, e.g. colon cancer, gastric cancer, sarcoma, lymphoma
 FT or leukemia
 XX
 PS Example 2; SEQ ID No 1601; 134pp; English.
 XX
 CC The invention describes a purified polypeptide, which comprises a
 CC fragment of a kinase, phosphatase, protease, protease inhibitor,
 CC transporter, cytoskeletal protein, receptor or transcription factor.
 CC The polypeptide is useful as an immunogenic composition for eliciting
 CC in a mammal an immunogenic response directed against any of the purified
 CC polypeptide. The purified polypeptide, or the antibody that binds to
 CC this polypeptide, is useful for treating cancer. The polypeptide is
 CC also useful for identifying compounds that binds to a naturally
 CC processed class I or class II MHC-binding polypeptide. The polypeptides
 CC and polynucleotides are particularly useful for treating or preventing
 CC myeloma, colon cancer, gastric cancer, adenocarcinoma, sarcoma, melanoma,
 CC lymphoma or leukaemia. These are also useful for screening agents for
 CC treating the above mentioned diseases. This sequence represents an
 CC expressed protein tag (EPT) isolated from human tissue for translational
 CC profiling.
 CC Note: This sequence does not appear in the printed specification but was
 CC obtained in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.
 XX
 XX Sequence 437 AA;

Query Match 99.8%; Score 2337; DB 24; Length 437;
 Best Local Similarity 100.0%; Pred. No. 9e-194;
 Matches 434; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 2 DPDSQPLNSLDVVKPLRPMPETFRKVGIPPIIIALLSLASIIIVVLIKVLIDKYFL 61
 Db 4 DPDSQPLNSLDVVKPLRPMPETFRKVGIPPIIIALLSLASIIIVVLIKVLIDKYFL 63
 QY 62 CGOPLHPIPRKQLCDGELDCPLGEDEEHCVKSPPEGPAVAVRLSKDRSTLQVLD SATGNW 121
 Db 64 CGOPLHPIPRKQLCDGELDCPLGEDEEHCVKSPPEGPAVAVRLSKDRSTLQVLD SATGNW 123
 QY 122 FSACFDNFTTEALAEACRQMGYSKPTFRAVEIGPDQDLVDWEITENSQELMRNSSGPC 181
 Db 124 FSACFDNFTTEALAEACRQMGYSKPTFRAVEIGPDQDLVDWEITENSQELMRNSSGPC 183
 QY 182 LSGSLVSLHCLACGKSLKTRPVVGGEEASVDSWPQVSIQYDKQHVCGSILDPHWWLTA 241
 Db 184 LSGSLVSLHCLACGKSLKTRPVVGGEEASVDSWPQVSIQYDKQHVCGSILDPHWWLTA 243
 QY 242 AHCFRKHDTVFNKVRAGSKLGSFSLAVAKIIIEFNPMYPKNDIALMKLQPLTFS 301
 Db 244 AHCFRKHDTVFNKVRAGSKLGSFSLAVAKIIIEFNPMYPKNDIALMKLQPLTFS 303
 QY 302 GTVRPCLPFDFBELTPATPLWIIIGWFTKQNGKMSDILLOASVQVIDSTRCNADDAYQ 361
 Db 304 GTVRPCLPFDFBELTPATPLWIIIGWFTKQNGKMSDILLOASVQVIDSTRCNADDAYQ 363

QY 362 GEVTEKMCAGIPGGVDTCQDGGGLMYQSDQHVHVGIVSWGCGGSPSTPGVYTKVS 421
Db 364 GEVTEKMCAGIPGGVDTCQDGGGLMYQSDQHVHVGIVSWGCGGSPSTPGVYTKVS 423
QY 422 AYLNWIYNNVKAEL 435
Db 424 AYLNWIYNNVKAEL 437

RESULT 11
ABU04936
ID ABU04936 standard; Protein; 437 AA.
AC ABU04936;
XX
XX 29-JAN-2003 (first entry)
XX Human expressed protein tag (EPT) #1602.
XX
XX Translational profiling; expressed protein tag; EPT; kinase;
KW phosphatase; protease; protease inhibitor; transporter;
KW cytoskeletal protein; receptor; transcription factor; cancer; MHC;
KW major histocompatibility complex; myeloma; colon cancer;
KW gastric cancer; adenocarcinoma; sarcoma; melanoma; lymphoma;
KW leukaemia.
XX
XX Homo sapiens.
OS
XX
XX WO200278524-A2.
XX
XX 10-OCT-2002.
XX
XX 28-MAR-2002; 2002WO-US09671.
XX
XX 28-MAR-2001; 2001US-279495P.
PR 21-MAY-2001; 2001US-292544P.
PR 08-AUG-2001; 2001US-310801P.
PR 01-OCT-2001; 2001US-326370P.
PR 04-DEC-2001; 2001US-336780P.
PR 20-FEB-2002; 2002US-358985P.
XX
XX (ZYCO-) ZYCOS INC.
XX
XX Chiciz RM, Tomlinson AJ, Urban RG;
PI
XX
XX WPI; 2003-040607/03.
XX
XX New polypeptides (e.g. kinases, phosphatases, proteases, transporters,
PT cytoskeletal proteins, receptors or transcription factors), useful for
PT treating cancer, e.g. colon cancer, gastric cancer, lymphoma
PT or leukemia -
XX
XX Example 2; SEQ ID No 1602; 134pp; English.
PS
XX The invention describes a purified polypeptide, which comprises a
XX fragment of a kinase, phosphatase, protease, protease inhibitor,
CC transporter, cytoskeletal protein, receptor or transcription factor.
CC The polypeptide is useful as an immunogenic composition for eliciting
CC in a mammal an immunogenic response directed against any of the purified
CC polypeptide. The purified polypeptide, or the antibody that binds to
CC this polypeptide, is useful for treating cancer. The polypeptide is
CC also useful for identifying compounds that binds to a naturally
CC processed class I or class II MHC-binding polypeptide. The polypeptides
CC and polynucleotides are particularly useful for treating or preventing
CC myeloma, colon cancer, gastric cancer, adenocarcinoma, sarcoma, melanoma,
CC lymphoma or leukaemia. These are also useful for screening agents for
CC treating the above mentioned diseases. This sequence represents an
CC expressed protein tag (EPT) isolated from human tissue for translational
CC profiling.
CC Note: This sequence does not appear in the printed specification but was
CC obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 437 AA;
Query Match 99.8%; Score 2337; DB 24; Length 437;
Best Local Similarity 100.0%; Pred. No. 9e-194;
Matches 434; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 DPDSDFPLNSLDVVKPLRKPIMPETFRKVGIPITIIALLSLASIIIVVVLKVLDKYFL 61
Db 4 DPDSDFPLNSLDVVKPLRKPIMPETFRKVGIPITIIALLSLASIIIVVVLKVLDKYFL 63
QY 62 CGOPLHFIPRKQLCDGELDCPLGDEBEHCVKSPPEGPAVAVRLSKORSTIQVLDSATGNW 121
Db 64 CGOPLHFIPRKQLCDGELDCPLGDEBEHCVKSPPEGPAVAVRLSKORSTIQVLDSATGNW 123
QY 122 FSACFDNFTEALAEATACROMGYSSKPTFRAVEIGPDQDLDDVEITENSQELRNRNSSGPC 181
Db 124 FSACFDNFTEALAEATACROMGYSSKPTFRAVEIGPDQDLDDVEITENSQELRNRNSSGPC 183
QY 182 LSGSLVSLHCLACGKSLKTPRVVVGEEASVDSWPQVSIQYDKQHVCGGSIILDPHWLTA 241
Db 184 LSGSLVSLHCLACGKSLKTPRVVVGEEASVDSWPQVSIQYDKQHVCGGSIILDPHWLTA 243
QY 242 AHCFRKHTDVFNWVKVAGSKLGSFSLAVAKIIIIIFENPMYKNDIALMKLOPFLTFS 301
Db 244 AHCFRKHTDVFNWVKVAGSKLGSFSLAVAKIIIIIFENPMYKNDIALMKLOPFLTFS 303
QY 302 GTVRPICLPFFDEELTPATPLMTIIGWFTKQNGKMSDIIILQASVOVIDSTRCNADDAYQ 361
Db 304 GTVRPICLPFFDEELTPATPLMTIIGWFTKQNGKMSDIIILQASVOVIDSTRCNADDAYQ 363
QY 362 GEVTEKMCAGIPGGVDTCQDGGGLMYQSDQHVHVGIVSWGCGGSPSTPGVYTKVS 421
Db 364 GEVTEKMCAGIPGGVDTCQDGGGLMYQSDQHVHVGIVSWGCGGSPSTPGVYTKVS 423
QY 422 AYLNWIYNNVKAEL 435
Db 424 AYLNWIYNNVKAEL 437

RESULT 12
AAY72559
ID AAY72559 standard; Protein; 492 AA.
XX
XX AAY72559;
XX
XX 02-MAY-2001 (first entry)
XX
XX Human seripancrin variant #1 protein.
XX
XX Human; seripancrin; serine protease; chromosome 11q22-q23; therapy;
KW arthritis; chronic obstructive pulmonary disorder; COPD; cancer;
KW osteoporosis; aberrant wound healing; angiogenesis; diabetes;
KW inflammatory disorder; stroke; cardiovascular disease; gene therapy;
KW vaccine; cytostatic; cerebroprotective; vulnery; osteopathic.
XX
XX Homo sapiens.
XX
XX WO200104141-A2.
XX
XX 18-JAN-2001.
XX
XX 04-JUL-2000; 2000WO-EP06211.
XX
XX 12-JUL-1999; 99EP-0113428.
XX
XX (MERE) MERCK PATENT GMBH.
XX
XX Suendermann B, Hofmann U, Matzku S, Wilbert O;
XX
XX WPI; 2001-147177/15.
XX
XX N-PSDB; AAD02557.
XX

PT New extracellular serine protease Seripancrin, useful for treating
PT cancer, osteoporosis, arthritis, chronic obstructive pulmonary disease,
PT diabetes, inflammatory disorders, stroke, angiogenesis and aberrant
PT wound healing -
XX PS Disclosure; Page 43-44; 45pp; English.
XX
CC The present invention relates to seripancrin polynucleotides, and
CC polypeptides encoded by them. Seripancrin are members of serine
CC protease family. This protein contains a transmembrane domain,
CC a low density lipoprotein (LDL) domain, protease domain and a
CC scavenger receptor cysteine-rich (SRCR) domain. The LDL and SRCR
CC domains help to define the specificity of seripancrin's intra and
CC intermolecular interactions. The polynucleotides and polypeptides
CC of the invention are useful for treating and diagnosing diseases
CC such as arthritis, chronic obstructive pulmonary disorder (COPD),
CC cancer, osteoporosis, aberrant wound healing, angiogenesis,
CC inflammatory disorders, diabetes, stroke and cardiovascular diseases.
CC Seripancrin genes are useful in chromosome localisation studies,
CC as tools for tissue expression studies and also in gene therapy.
CC The polypeptides of the invention are used for identifying agonists
CC and antagonists useful for treating conditions associated with
CC seripancrin imbalance. These polypeptides are also useful as vaccines.
CC The present sequence is seripancrin variant #1 protein. The
CC seripancrin gene is located on human chromosome 11q22-q23.
XX
XX Sequence 492 AA;
XX
XX Query Match 99.4%; Score 2329; DB 22; Length 492;
XX Best Local Similarity 100.0%; Pred. No. 5.1e-193;
XX Matches 432; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MDPDSQPLNSLDVVKLPKPRIPMETFRKVGIPITIALSLASIIIVVLIKVLIDKYYF 60
DB 1 MDPDSQPLNSLDVVKLPKPRIPMETFRKVGIPITIALSLASIIIVVLIKVLIDKYYF 60
QY 61 LCGQPLHPIPRKQICDGLDCPLGEDEEHCVKSPFEGPAVAVRLSKDRSTLQVLDSATGN 120
DB 61 LCGQPLHPIPRKQICDGLDCPLGEDEEHCVKSPFEGPAVAVRLSKDRSTLQVLDSATGN 120
QY 121 WFSACPDNFTALAEATACRQMGYSKPTFRVBEIGPDQDLVDVVEITENSQELMRNSSGP 180
DB 121 WFSACPDNFTALAEATACRQMGYSKPTFRVBEIGPDQDLVDVVEITENSQELMRNSSGP 180
QY 181 CLSGSLVSLHCLACGSLKTPRVVGGEEASVDSWPQVSIQYDKQVCGGSLDPHWVLT 240
DB 181 CLSGSLVSLHCLACGSLKTPRVVGGEEASVDSWPQVSIQYDKQVCGGSLDPHWVLT 240
QY 241 AAHCFRKHDTDFNWKVRAGSDKLGSPFSLAVAKIIIEFNPMYPKNDIALMKLQPLTF 300
DB 241 AAHCFRKHDTDFNWKVRAGSDKLGSPFSLAVAKIIIEFNPMYPKNDIALMKLQPLTF 300
QY 301 SGTVRPCLPFDDELTPTPLWIIWGFTKQNGKMSDILLQASVQVIDSTRCNADDAY 360
DB 301 SGTVRPCLPFDDELTPTPLWIIWGFTKQNGKMSDILLQASVQVIDSTRCNADDAY 360
QY 361 QGEVTEKMKMAGIPEGGVDTCCGDSGLMYOSDQWHVGVISWGYCGGSPTPGYTKV 420
DB 361 QGEVTEKMKMAGIPEGGVDTCCGDSGLMYOSDQWHVGVISWGYCGGSPTPGYTKV 420
QY 421 SAYLNWIYNVWK 432
DB 421 SAYLNWIYNVWK 432
RESULT 13
ABU04913
ID ABU04913 standard; Protein; 492 AA.
XX AC
XX AC ABU04913;
XX
XX 29-JAN-2003 (first entry)
XX DT
XX

DE Human expressed protein tag (EPT) #1579.
XX
KW Translational profiling; expressed protein tag; EPT; kinase;
KW phosphatase; protease; protease inhibitor; transporter;
KW cytoskeletal protein; receptor; transcription factor; cancer; MHC;
KW major histocompatibility complex; myeloma; colon cancer;
KW gastric cancer; adenocarcinoma; sarcoma; melanoma; lymphoma;
KW leukaemia.
XX
OS Homo sapiens.
XX
DN WO200278524-A2.
XX
PD 10-OCT-2002.
XX
PF 28-MAR-2002; 2002WO-US09671.
XX
PR 28-MAR-2001; 2001US-279495P.
PR 21-MAY-2001; 2001US-292544P.
PR 08-AUG-2001; 2001US-310801P.
PR 01-OCT-2001; 2001US-326370P.
PR 04-DEC-2001; 2001US-336780P.
PR 20-FEB-2002; 2002US-358985P.
XX
XX (ZYCO-) ZYCO INC.
XX
XX Chiciz RM, Tomlinson AJ, Urban RG;
XX WPI; 2003-040607/03.
XX
XX New polypeptides (e.g. kinases, phosphatases, proteases, transporters,
XX cytoskeletal proteins, receptors or transcription factors), useful for
XX treating cancer, e.g. colon cancer, gastric cancer, sarcoma, lymphoma
XX or leukemia -
XX
XX Example 2; SEQ ID No 1579; 134pp; English.
XX
XX The invention describes a purified polypeptide, which comprises a
XX fragment of a kinase, phosphatase, protease, protease inhibitor,
XX transporter, cytoskeletal protein, receptor or transcription factor.
XX The polypeptide is useful as an immunogenic composition for eliciting
XX in a mammal an immunogenic response directed against any of the purified
XX polypeptide. The purified polypeptide, or the antibody that binds to
XX this polypeptide, is useful for treating cancer. The polypeptide is
XX also useful for identifying compounds that binds to a naturally
XX processed class I or class II MHC-binding polypeptide. The polypeptides
XX and polynucleotides are particularly useful for treating or preventing
XX myeloma, colon cancer, gastric cancer, adenocarcinoma, sarcoma, melanoma,
XX lymphoma or leukaemia. These are also useful for screening agents for
XX treating the above mentioned diseases. This sequence represents an
XX expressed protein tag (EPT) isolated from human tissue for translational
XX profiling.
XX Note: This sequence does not appear in the printed specification but was
XX obtained in electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 492 AA;
XX
XX Query Match 99.4%; Score 2329; DB 24; Length 492;
XX Best Local Similarity 100.0%; Pred. No. 5.1e-193;
XX Matches 432; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MDPDSQPLNSLDVVKLPKPRIPMETFRKVGIPITIALSLASIIIVVLIKVLIDKYYF 60
DB 1 MDPDSQPLNSLDVVKLPKPRIPMETFRKVGIPITIALSLASIIIVVLIKVLIDKYYF 60
QY 61 LCGQPLHPIPRKQICDGLDCPLGEDEEHCVKSPFEGPAVAVRLSKDRSTLQVLDSATGN 120
DB 61 LCGQPLHPIPRKQICDGLDCPLGEDEEHCVKSPFEGPAVAVRLSKDRSTLQVLDSATGN 120
QY 121 WFSACPDNFTALAEATACRQMGYSKPTFRVBEIGPDQDLVDVVEITENSQELMRNSSGP 180
DB 121 WFSACPDNFTALAEATACRQMGYSKPTFRVBEIGPDQDLVDVVEITENSQELMRNSSGP 180

181 CLSGSLVSLHCLACGSLKTPRVVGGEEASVDSWPQVSIQYDKQHVCGSILDPHWVLT 240
181 CUSGSLVSLHCLACGSLKTPRVVGGEEASVDSWPQVSIQYDKQHVCGSILDPHWVLT 240
241 AAHCFRKHDTVFNWVKRAGSDKLGSPSLAVAKIIIEFNPMYPKNDIALMKLOPPLTF 300
241 AAHCFRKHDTVFNWVKRAGSDKLGSPSLAVAKIIIEFNPMYPKNDIALMKLOPPLTF 300
301 SGTVPICLPFFDEELTPATPLWIIIGWGTCKQNGKMSDILLQASVQVIDSTRCNADDAY 360
301 SGTVPICLPFFDEELTPATPLWIIIGWGTCKQNGKMSDILLQASVQVIDSTRCNADDAY 360
361 QGEVTEKMKACGIPGEGVDTCCGDSGGLMYQSDQVHVVGVISWGYCGGPGSTPGVYTKV 420
361 QGEVTEKMKACGIPGEGVDTCCGDSGGLMYQSDQVHVVGVISWGYCGGPGSTPGVYTKV 420
421 SAYLNWIYNVWK 432
421 SAYLNWIYNVWK 432

RESULT 14
ABU04932
ID ABU04932 standard; Protein; 492 AA.
XX AC ABU04932;
XX AC ABU04932;
XX DT 29-JAN-2003 (first entry)
XX DT Human expressed protein tag (EPT) #1598.
XX XX Translational profiling; expressed protein tag; EPT; kinase;
XX KW phosphatase; protease; protease inhibitor; transporter;
XX KW cytoskeletal protein; receptor; transcription factor; cancer; MHC;
XX KW major histocompatibility complex; myeloma; colon cancer;
XX KW gastric cancer; adenocarcinoma; sarcoma; melanoma; lymphoma;
XX KW leukaemia.
XX CS Homo sapiens.
XX CS WO200278524-A2.
XX XX 10-OCT-2002.
XX XX 28-MAR-2002; 2002WO-US09671.
XX XX 28-MAR-2001; 2001US-279495P.
XX XX 21-MAY-2001; 2001US-292544P.
XX XX 08-AUG-2001; 2001US-310801P.
XX XX 01-OCT-2001; 2001US-326370P.
XX XX 04-DEC-2001; 2001US-336780P.
XX XX 20-FEB-2002; 2002US-358985P.
XX XX (ZYCO-) ZYCOS INC.
XX XX Chicx RM, Tomlinson AJ, Urban RG;
XX XX WPI; 2003-040607/03.
XX DR New polypeptides (e.g. kinases, phosphatases, proteases, transporters,
XX PT cytoskeletal proteins, receptors or transcription factors), useful for
XX PT treating cancer, e.g. colon cancer, gastric cancer, sarcoma, lymphoma
XX PT or leukemia
XX PS Example 2; SEQ ID No 1598; 134pp; English.
XX PS The invention describes a purified polypeptide, which comprises a
XX PS fragment of a kinase, phosphatase, protease, protease inhibitor,
XX PS transporter, cytoskeletal protein, receptor or transcription factor.
XX PS The polypeptide is useful as an immunogenic composition for eliciting
XX PS in a mammal an immunogenic response directed against any of the purified
XX PS polypeptide. The purified polypeptide, or the antibody that binds to

CC this polypeptide, is useful for treating cancer. The polypeptide is
CC also useful for identifying compounds that binds to a naturally
CC processed class I or class II MHC-binding polypeptide. The polypeptides
CC and polynucleotides are particularly useful for treating or preventing
CC myeloma, colon cancer, gastric cancer, adenocarcinoma, sarcoma, melanoma,
CC lymphoma or leukaemia. These are also useful for screening agents for
CC treating the above mentioned diseases. This sequence represents an
CC expressed protein tag (EPT) isolated from human tissue for translational
CC profiling.
CC Note: This sequence does not appear in the printed specification but was
CC obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX XX
SQ Sequence 492 AA;
Query Match 99.4%; Score 2329; DB 24; Length 492;
Best Local Similarity 100.0%; Pred. No. 5.1e-193;
Matches 432; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MDPDSQPLNSLDVVKPLKPRIPMETFRKVGIPPIIALLSLASIIIVVVLKIVLDKIYF 60
DB 1 MDPDSQPLNSLDVVKPLKPRIPMETFRKVGIPPIIALLSLASIIIVVVLKIVLDKIYF 60
QY 61 LCGQPLHPIPRKQLDGLDPLGDEDEHCVKSPPEGPAVAVRLSKDRSTLQVLDSATGN 120
DB 61 LCGQPLHPIPRKQLDGLDPLGDEDEHCVKSPPEGPAVAVRLSKDRSTLQVLDSATGN 120
QY 121 WFSACFDNFTALAEATACRQNGYSKPTFRAVEIGDQDLDVVEITENSQELRMNNSGP 180
DB 121 WFSACFDNFTALAEATACRQNGYSKPTFRAVEIGDQDLDVVEITENSQELRMNNSGP 180
QY 181 CLSGSLVSLHCLACGSLKTPRVVGGEEASVDSWPQVSIQYDKQHVCGSILDPHWVLT 240
DB 181 CLSGSLVSLHCLACGSLKTPRVVGGEEASVDSWPQVSIQYDKQHVCGSILDPHWVLT 240
QY 241 AAHCFRKHDTVFNWVKRAGSDKLGSPSLAVAKIIIEFNPMYPKNDIALMKLOPPLTF 300
DB 241 AAHCFRKHDTVFNWVKRAGSDKLGSPSLAVAKIIIEFNPMYPKNDIALMKLOPPLTF 300
QY 301 SGTVPICLPFFDEELTPATPLWIIIGWGTCKQNGKMSDILLQASVQVIDSTRCNADDAY 360
DB 301 SGTVPICLPFFDEELTPATPLWIIIGWGTCKQNGKMSDILLQASVQVIDSTRCNADDAY 360
QY 361 QGEVTEKMKACGIPGEGVDTCCGDSGGLMYQSDQVHVVGVISWGYCGGPGSTPGVYTKV 420
DB 361 QGEVTEKMKACGIPGEGVDTCCGDSGGLMYQSDQVHVVGVISWGYCGGPGSTPGVYTKV 420
QY 421 SAYLNWIYNVWK 432
DB 421 SAYLNWIYNVWK 432

RESULT 15
AAE06931
ID AAE06931 standard; Protein; 437 AA.
XX AC AAE06931;
XX DT 16-OCT-2001 (first entry)
XX DT Human membrane-type serine protease (MTSP) 3.
XX DE Human; transmembrane serine protease; membrane-type serine protease;
XX KW MTSP; protease domain; neoplastic disease; tumour; cancer; cytostatic;
XX KW lung carcinoma; colon adenocarcinoma; ovarian carcinoma; gene therapy.
XX OS Homo sapiens.
XX XX
XX FH Key Location/Qualifiers
XX FT Domain 205..437
XX FT /label= Protease_domain
XX FT Cleavage-site 204..205
XX FT Misc-difference 310

/note= "Unpaired cysteine"

FT XX

FN PN

PD PD

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

WO200157194-A2.

09-AUG-2001.

02-FEB-2001; 2001WO-US03471.

03-FEB-2000; 2000US-0179982.

18-FEB-2000; 2000US-0183542.

22-JUN-2000; 2000US-0213124.

26-JUL-2000; 2000US-0220970.

08-SEP-2000; 2000US-0657986.

22-SEP-2000; 2000US-0234840.

(CORV-) CORVAS INT INC.

Madison EL, Ong EO, Yeh J;

WPI; 2001-488877/53.

N-PSDB; AADI3114.

Novel single chain polypeptide comprising protease domain of type-II

membrane-type serine protease or its catalytically active portion

useful for treating and preventing cancer and tumor

Claim 12; Page 199-200; 256pp; English.

The invention relates to transmembrane serine proteases and their

corresponding nucleotides and the protease domain of a type-II

membrane-type serine protease (MTSP). MTSP is useful for identifying

compounds that modulate or inhibit its proteolytic activity and for

formulating a medicament for treating neoplastic disease. MTSP and

its corresponding nucleotides are useful in preventing or treating

tumors or cancers such as lung carcinoma, colon adenocarcinoma and

ovarian carcinoma, in diagnostics and in hybridization assays. MTSP

is useful as a diagnostic marker for tumor development, growth and/or

progression and as immunogens to generate antibodies that specifically

bind to it. MTSP DNA is useful in a yeast two-hybrid system and in

gene therapy. The present sequence is human MTSP3 protein.

Sequence 437 AA;

Query Match

Best Local Similarity 99.0%; Score 2319; DB 22; Length 437;

Matches 432; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

2 DPDSQPLNSLDVVKPKPRIEMETPRKVGIPITIIALLSLASIIIVVLKVIDKYFL 61

4 DPDSQPLNSLDVVKPKPRIEMETPRKVGIPITIIALLSLASIIIVVLKVIDKYFL 63

62 CQOPLHFIIPKOLCDGELDCPLGDEEHCHVKSPFPGPAVAVRLSKDRSTLOVLSATGNW 121

64 CQOPLHFIIPKOLCDGELDCPLGDEEHCHVKSPFPGPAVAVRLSKDRSTLOVLSATGNW 123

122 PSACFDNFTALAEATACROMGYSSKPTFAVEIGDDDLVDVVEITENSQELMRNSSGPC 181

124 PSACFDNFTALAEATACROMGYSSKPTFAVEIGDDDLVDVVEITENSQELMRNSSGPC 183

182 LSGSLVSLHCLACGSLKTPRVVGGEEASVDSWPMQVSIQYDKQHVCGGSILDPHWLTA 241

184 LSGSLVSLHCLACGSLKTPRVVGGEEASVDSWPMQVSIQYDKQHVCGGSILDPHWLTA 243

242 AHCFRKHDTDFNWKVRAGSKLGSPFSLAVAKIIIEFNPMYPKNDIALMKLOPPLTFS 301

244 AHCFRKHDTDFNWKVRAGSKLGSPFSLAVAKIIIEFNPMYPKNDIALMKLOPPLTFS 303

302 GTVREICLPFFDEELTPATPLIIGWFTKQNGKMSDILLQASVQVIDSTRCNADDAYQ 361

304 GTVREICLPFFDEELTPATPLIIGWFTKQNGKMSDILLQASVQVIDSTRCNADDAYQ 363

362 GEVTEKMMCAGIPEGGVDTCCQDSGGPLMYQSDQHWVGIWSWGYCGGPGSTPGVYTKVS 421

422 AYLNWIYNWKAEEL 435

424 AYLNWIYNWKAEEL 437

Search completed: November 23, 2003, 07:49:05

Job time : 68.0234 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 23, 2003, 07:47:10 ; Search time 22.1389 Seconds
(without alignments)
831.351 Million cell updates/sec

Title: US-09-607-745-2

Perfect score: 2342

Sequence: 1 WDPDSQPLNSLDVPLRKP.....VYTKVSAIYNWYVKAEL 435

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:*

1: /cgn2_6/prodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/prodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/prodata/2/iaa/5A_COMB.pep:*
4: /cgn2_6/prodata/2/iaa/5B_COMB.pep:*
5: /cgn2_6/prodata/2/iaa/PCUTS_COMB.pep:*
6: /cgn2_6/prodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2338	99.8	435	3	US-09-008-271A-6
2	2196	93.8	423	4	US-09-656-002-2
3	676.5	28.9	492	3	US-09-342-749-2
4	676.5	28.9	492	4	US-09-691-840-2
5	655.5	28.0	454	3	US-09-518-046-2
6	588.5	25.1	417	4	US-09-820-002-4
7	584	24.9	455	3	US-09-261-416-2
8	580	24.8	376	4	US-09-820-002-2
9	574	24.5	416	2	US-09-000-846-2
10	571	24.4	798	5	PCN-US94-00616-2
11	571	24.4	798	5	PCN-US94-00616-2
12	568	24.3	418	1	US-08-508-448C-25
13	568	24.3	418	4	US-09-370-838-82
14	568	24.3	418	4	US-09-370-838-82
15	566	24.2	418	4	US-09-370-838-82
16	558.5	23.8	283	3	US-08-807-151-1
17	558.5	23.8	283	4	US-09-478-957-1
18	531.5	22.7	232	1	US-08-508-448C-19
19	512.5	21.9	256	2	US-09-027-337-3
20	512.5	21.9	256	4	US-09-644-600-3
21	511	21.8	638	2	US-08-681-151-3
22	507.5	21.7	255	3	US-08-944-483-67
23	503.5	21.5	285	4	US-09-023-942A-26
24	486	20.8	248	3	US-08-944-483-63
25	483	20.6	314	4	US-09-023-942A-6
26	482.5	20.6	284	4	US-09-387-375-7
27	478	20.4	314	3	US-09-008-271A-3

28 477 20.4 312 4 US-09-023-942A-4 Sequence 4, Appli
29 475 20.3 407 4 US-09-734-675-4 Sequence 4, Appli
30 472.5 20.2 317 4 US-09-386-629-7 Sequence 7, Appli
31 470 20.1 902 4 US-09-644-600-10 Sequence 10, Appli
32 467.5 20.0 290 4 US-09-386-653A-7 Sequence 7, Appli
33 466.5 19.9 405 4 US-09-734-675-2 Sequence 2, Appli
34 465.5 19.9 250 3 US-08-944-483-68 Sequence 68, Appli
35 463 19.8 855 2 US-09-027-337-2 Sequence 2, Appli
36 463 19.8 855 4 US-09-644-600-2 Sequence 2, Appli
37 458.5 19.6 316 4 US-09-387-375-9 Sequence 9, Appli
38 458.5 19.6 790 4 US-08-991-761A-13 Sequence 13, Appli
39 455.5 19.4 812 1 US-08-248-629A-1 Sequence 1, Appli
40 455.5 19.4 812 1 US-08-451-932-1 Sequence 1, Appli
41 455.5 19.4 812 1 US-08-452-260-1 Sequence 1, Appli
42 455.5 19.4 812 1 US-08-326-785-1 Sequence 1, Appli
43 455.5 19.4 812 2 US-08-612-788-1 Sequence 1, Appli
44 455.5 19.4 812 2 US-08-605-598B-1 Sequence 1, Appli
45 455.5 19.4 812 2 US-08-429-743-1 Sequence 1, Appli

ALIGNMENTS

RESULT 1

US-09-008-271A-6

; Sequence 6, Application US/09008271A

; Patent No. 6203979

; GENERAL INFORMATION:

; APPLICANT: Bandman, Olga

; Yue, Henry

; Guegler, Karl J.

; Corley, Neil C.

; Tang, Tom Y.

; Shah, Purvi

; TITLE OF INVENTION: HUMAN PROTEASE MOLECULES

; NUMBER OF SEQUENCES: 24

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Incyte Pharmaceuticals, Inc.

; STREET: 3174 Porter Dr.

; CITY: Palo Alto

; STATE: CA

; COUNTRY: USA

; ZIP: 94304

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; SOFTWARE: FastSeq for Windows Version 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/008,271A

; FILING DATE: 16-Jan-1998

; PRIORITY APPLICATION DATA:

; FILING DATE: <Unknown>

; ATTORNEY/AGENT INFORMATION:

; NAME: Mohan-Peterson, Sheela

; REGISTRATION NUMBER: 41,201

; REFERENCE/DOCKET NUMBER: PF-0458 US

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 650-855-0555

; TELEFAX: 650-845-4166

; INFORMATION FOR SEQ ID NO: 6:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 435 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; IMMEDIATE SOURCE:

; LIBRARY: COLNNOT13

; CLONE: 1337018

; SEQUENCE DESCRIPTION: SEQ ID NO: 6 :

US-09-008-271A-6

Query Match 99.8%; Score 2338; DB 3; Length 435;
Best Local Similarity 99.8%; Pred. No. 6.2e-231;
Matches 434; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDPSDPLNSLDVVKPRKPRIPMETPRKVGIPITIIALLSLASIIIVVLIKVILDKYIF 60
DB 1 MDPSDPLNSLDVVKPRKPRIPMETPRKVGIPITIIALLSLASIIIVVLIKVILDKYIF 60

QY 61 LCQGPLHFIIPRKQCDGELDCPLGEDEBEHCVKSPPEGPAVAVRLSKDRSTLOVLDSATGN 120
DB 61 LCQGPLHFIIPRKQCDGELDCPLGEDEBEHCVKSPPEGPAVAVRLSKDRSTLOVLDSATGN 120

QY 121 WFSACPNFTEALAEACRQMGYSKPTTFRAVEIGPDQDLVVEITENSQELMRNSSGP 180
DB 121 WFSACPNFTEALAEACRQMGYSKPTTFRAVEIGPDQDLVVEITENSQELMRNSSGP 180

QY 181 CLSGSLVSLHCLACGSLKTPRVVVGGEASVDSWPMQVSIQYDKQHVCGGSIILDPHWLT 240
DB 181 CLSGSLVSLHCLACGSLKTPRVVVGGEASVDSWPMQVSIQYDKQHVCGGSIILDPHWLT 240

QY 241 AAHCPRKHTDVFNWKVRAGSKLGSFPLSAVAKIIIEFNPMYPKNDIALMKLOFPLTF 300
DB 241 AAHCPRKHTDVFNWKVRAGSKLGSFPLSAVAKIIIEFNPMYPKNDIALMKLOFPLTF 300

QY 301 SGTVPICLPFFDEELTPATPLMIIGWFTKONGKMSDILLQASVOVIDSTRCNADDDAY 360
DB 301 SGTVPICLPFFDEELTPATPLMIIGWFTKONGKMSDILLQASVOVIDSTRCNADDDAY 360

QY 361 QGEVTEKMCAGIPEGGVDTFCQDSDGSGPLMYQSDQHVGVGIVSGYCGGSPSTPGVYTKV 420
DB 361 QGEVTEKMCAGIPEGGVDTFCQDSDGSGPLMYQSDQHVGVGIVSGYCGGSPSTPGVYTKV 420

QY 421 SAYLNWIYNWKAE 435
DB 421 SAYLNWIYNWKAE 435

RESULT 2
US-09-656-002-2
; Sequence 2, Application US/09656002
; Patent No. 6455668
; GENERAL INFORMATION:
; APPLICANT: Mack, David
; APPLICANT: Gish, Kurt
; APPLICANT: Wilson, Keith
; TITLE OF INVENTION: NOVEL METHODS OF DIAGNOSING COLORECTAL CANCER, COMPOSITIONS, AND
; FILE REFERENCE: A-69108/DJB/JUD/AMS
; CURRENT APPLICATION NUMBER: US/09/656,002
; CURRENT FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: US 09/525,993
; PRIOR FILING DATE: 2000-03-15
; PRIOR APPLICATION NUMBER: US 09/493,444
; PRIOR FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: PCT/US 00/07044
; PRIOR FILING DATE: 2000-03-15
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 2
; LENGTH: 423
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-656-002-2

Query Match 93.8%; Score 2196; DB 4; Length 423;
Best Local Similarity 97.1%; Pred. No. 2.1e-216;
Matches 408; Conservative 1; Mismatches 11; Indels 0; Gaps 0;

QY 16 PRKPRIPMETPRKVGIPITIIALLSLASIIIVVLIKVILDKYIFLCQGPLHFIIPRKQIC 75
DB 4 PCANFVSPWRPSESIGVPIITIIALLSLASIIIVVLIKVILDKYIFLCQGPLHFIIPRKQIC 63

QY 76 DGELDCPLGEDEBEHCVKSPPEGPAVAVRLSKDRSTLOVLDSATGNWFSACFDNFTALAE 135
DB 64 DGELDCPLGEDEBEHCVKSPPEGPAVAVRLSKDRSTLOVLDSATGNWFSACFDNFTALAE 123

QY 136 TACRQMGYSKPTTFRAVEIGPDQDLVVEITENSQELMRNSSGPCLSGSLVSLHCLACG 195
DB 124 TACRQMGYSKPTTFRAVEIGPDQDLVVEITENSQELMRNSSGPCLSGSLVSLHCLACG 183

QY 196 KSLKTPRVVVGGEASVDSWPMQVSIQYDKQHVCGGSIILDPHWLTAACHFRKHTDVFNWK 255
DB 184 KSLKTPRVVVGGEASVDSWPMQVSIQYDKQHVCGGSIILDPHWLTAACHFRKHTDVFNWK 243

QY 256 VRAGSKLGSFPLSAVAKIIIEFNPMYPKNDIALMKLOFPLTFSGTVRPICLPFPDEE 315
DB 244 VRAGSKLGSFPLSAVAKIIIEFNPMYPKNDIALMKLOFPLTFSGTVRPICLPFPDEE 303

QY 316 LTPATPLMIIGWFTKONGKMSDILLQASVOVIDSTRCNADDDAYQGEVTEKMCAGIPE 375
DB 304 LTPATPLMIIGWFTKONGKMSDILLQASVOVIDSTRCNADDDAYQGEVTEKMCAGIPE 363

QY 376 GGVDTFCQDSDGSGPLMYQSDQHVGVGIVSGYCGGSPSTPGVYTKVSAYLNIYNWKAE 435
DB 364 GGVDTFCQDSDGSGPLMYQSDQHVGVGIVSGYCGGSPSTPGVYTKVSAYLNIYNWKAE 423

RESULT 3
US-09-342-749-2
; Sequence 2, Application US/09342749
; Patent No. 6166194
; GENERAL INFORMATION:
; APPLICANT: Wong, Alexander K.C.
; APPLICANT: Tavtigian, Sean V.
; APPLICANT: Teng, David H.-F.
; APPLICANT: Myriad Genetics, Inc.
; TITLE OF INVENTION: TSPRS2 is a Tumor Suppressor
; FILE REFERENCE: 2318-202
; CURRENT APPLICATION NUMBER: US/09/342,749
; CURRENT FILING DATE: 1999-06-29
; EARLIER APPLICATION NUMBER: US 60/091,044
; EARLIER FILING DATE: 1998-06-29
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 2
; LENGTH: 492
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-342-749-2

Query Match 28.9%; Score 676.5; DB 3; Length 492;
Best Local Similarity 39.1%; Pred. No. 1.2e-60;
Matches 150; Conservative 57; Mismatches 128; Indels 49; Gaps 13;

QY 75 CDGELDCPLGEDEBEHCVKSPPEGPAVAVRLSKDRSTLOVLDSATGNWFSACFDNFTALAE 134
DB 133 CDGSHCHPGGDEENRCVLY--GP-----NFILQVSSQRKSWHPVQDDWNNYNG 181

QY 135 ETACRQMGYSKPTTFRAVEIGPDQDLVVEITENSQELMRNSSG-----PCL 182
DB 182 RAACEDMCGY--KNMFYSSQ-----GIVDSGSGTSFKMLNTSAGNVDIYKLYHSDACS 232

QY 183 SGSLVSLHCLACGKSL---KTPRVVVGGEASVDSWPMQVSIQYDKQHVCGGSIILDPHWL 239
DB 233 SKAVVSLRACIACGVNLNSSRQSRIVGGSALPGAWPQVSLHVQNVHVCGGSIITPEWIV 292

QY 240 TAACHFRKH-TDVFNWKVRAGSKLGSF-----PSLAVAKIIIEFNPMY-----PKNDIAL 291
DB 293 TAACHFRKPLNPNWHTAFAGILR-QSFMFYGAGYQVEKVI---SHPNYDSKTKNDIAL 348

QY 292 MKLOFPLTFSGTVRPICLPFPDEELTPATPLMIIGWFTKONGKMSDILLQASVOVIDS 351
DB 349 MKLOFPLTFDLNFKPCLPNFMNMQPEQLCWISGWGATEBK-GKTSEVLNAAKVLLET 407

QY 352 TRCNADDDAYQGEVTEKMCAGIPEGGVDTFCQDSDGSGPLMYQSDQ--WHVVGIVSGYCGG 410

Db 408 QRCNSRVYVNDLITPAMICAGFLQGNVDSQGDGGPLVTSKNIWLLIGDTSWGSCAK 467
QY 411 PSTPGVYTKVSAVNLWYVNWKAEL 434
Db 468 AYRPGVYGNVWFTDWIYQMRAD 491

RESULT 4

US-09-691-840-2
; Sequence 2, Application US/09691840
; Patent No. 6444419
; GENERAL INFORMATION:
; APPLICANT: Wong, Alexander K.C.
; APPLICANT: Tavtigian, Sean V.
; APPLICANT: Teng, David H.-F.
; APPLICANT: Myriad Genetics, Inc.
; TITLE OF INVENTION: TMPRSS2 is a Tumor Suppressor
; FILE REFERENCE: 2318-202
; CURRENT APPLICATION NUMBER: US/09/691,840
; CURRENT FILING DATE: 2000-10-18
; PRIOR APPLICATION NUMBER: US/09/342,749
; PRIOR FILING DATE: 1999-06-29
; PRIOR APPLICATION NUMBER: US 60/091,044
; PRIOR FILING DATE: 1998-06-29
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 2
; LENGTH: 492
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-691-840-2

Query Match 28.9%; Score 676.5; DB 4; Length 492;
Best Local Similarity 39.3%; Pred. No. 1.2e-60;
Matches 150; Conservative 57; Mismatches 128; Indels 49; Gaps 13;

QY 75 CDGELDCPLGDEDEHCVKSPPEGPAVAVRLSKDRSTLQVLD SATGNWFSACDFNFTAL 134
Db 133 CDGVSHCPGDEENRCURLY--GP-----NFIQVSSORKSWHPVQCDDWNNY 181
QY 135 ETACRQMGYSKPTFRAVEIGPDQDLVDVEITENSQELRNSSG-----PCL 182
Db 182 RAACRDWGY--KNNFYSSQ-----GIVDDSGSTSPMKLNTSAGNVDIYKLYHSDACS 232
QY 183 SGLSVSLHCLACGKSL---KTPRVGGEASVDSWPQVSIQYDKQHVCGGSIIDPHWVL 239
Db 233 SKAVSLRCITACGNLNSRQSRIVGESALPGAWPQVSLHVQNVLHVCGGSIITPEWIV 292
QY 240 TAAHCFRKH--TDVFNKVRAGSKLGSF---PSLAVAKIIIEFNPMY---PKNDIAL 291
Db 293 TAAHCVKEKPLNPNWHTAFAGILR--QSFMYGAGYQVEKVI---SHPNVDSTKKNNDIAL 348
QY 292 MKLQPLTSGTVRPICLPFDFDELTPATPLWILGWFTKQNGKMSDILLOASVQVIDS 351
Db 349 MKLQKPLTNDLVKPVCLPNPGMWLQPEQLCWLISGNGATEEK--GKTSEVLNAAKVLIIET 407
QY 352 TSCNADDAVQGEVTEKMKCAGIPPEGVDTCQDGGPLMYQSDQ--WHVVGIVSWGVCGG 410
Db 408 QRCNSRVYVNDLITPAMICAGFLQGNVDSQGDGGPLVTSKNIWLLIGDTSWGSCAK 467
QY 411 PSTPGVYTKVSAVNLWYVNWKAEL 434
Db 468 AYRPGVYGNVWFTDWIYQMRAD 491

RESULT 5

US-09-518-046-2
; Sequence 2, Application US/09518046
; Patent No. 6294663
; GENERAL INFORMATION:
; APPLICANT: O'Brien, Timothy J.
; APPLICANT: Underwood, Lowell J.

; TITLE OF INVENTION: Transmembrane Serine Protease Overexpressed
; TITLE OF INVENTION: in Ovarian Carcinoma and Uses Thereof
; FILE REFERENCE: D6192CIP
; CURRENT APPLICATION NUMBER: US/09/518,046
; CURRENT FILING DATE: 2000-03-02
; EARLIER APPLICATION NUMBER: 09/261,416
; EARLIER FILING DATE: 1999-03-03
; NUMBER OF SEQ ID NOS: 153
; SEQ ID NO 2
; LENGTH: 454
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: complete amino acid sequence of TADG-12
; OTHER INFORMATION: protein
US-09-518-046-2

Query Match 28.0%; Score 655.5; DB 3; Length 454;
Best Local Similarity 35.7%; Pred. No. 1.5e-58;
Matches 158; Conservative 73; Mismatches 149; Indels 63; Gaps 18;

QY 22 IPMETFRKVGIPFIILALLSLASIIIVVLIKVILD---KY-----YFLCGQPLHFIPRQQL 74
Db 44 LPPEVFSQSSSLGIILAL-----ILALAIGLGHFDGSGKYRCRSSFKC---IELITR--- 92
QY 75 CDGELDCPLGDEDEHCVKSPPEGPAVAVRLSKDRSTLQVLD SATGNWFSACDFNFTAL 134
Db 93 CDGVSDCKDGEDERYC-----VRVGQNAVLVQVFTAA--SWKTCSDDMKGHYA 139
QY 135 ETACRQMGYSK-----KPTFRAVEIGPDQDLVDVEITENSQELRNSSGPPCL 182
Db 140 NVACAGLGFPSVSSDNLVSLGQFREFEYIDHLLPDDKVTAHLHSHVYVREG---CA 196
QY 183 SGLSVSLHCLACG--KSLKTPRVVGGEEASVDSWPQVSIQYDKQHVCGGSIIDPHWVLTA 241
Db 197 SGHVTLQCTACGHRGYSRRIVGNNMSSLQWPMQASLQFGYHLGCGSVITPLWITA 256
QY 242 ANCFRKHDTVF---NWKVPAGSDKL--GSFPLAVAKIIIEFNPMY--PK--DNDIALMK 293
Db 257 ABCV---YDLVLPKNTIQVGLVSLDNPAPSHLVEKIV---YHSKYKPKRLGNDIALMK 310
QY 294 LQPLTFSGTVRPICLPFDFDELTPATPLWILGWFTKQNGKMSDILLOASVQVIDSTR 353
Db 311 LAGPLTFNEMIQVCLPNSSENFPGKVCWTSGWAT--EDGGDASVNLNHAAPVPLISNKI 369
QY 354 CNADDAVQGEVTEKMKCAGIPPEGVDTCQDGGPLMYQSDQ--WHVVGIVSWGVCGGPS 412
Db 370 CNHRDVYGGIISPSMLCAGYLTGGVDSQGDGGPLVCCQERLWKLVGATSPGIGCAEVN 429
QY 413 TPGVYTKVSAVNLWYVNWKAEL 435
Db 430 KEGVYTRVTSFLDNIHQMERDL 452

RESULT 6

US-09-820-002-4
; Sequence 4, Application US/09820002
; Patent No. 6482630
; GENERAL INFORMATION:
; APPLICANT: Gan, Weinui
; APPLICANT: Ye, Jane
; APPLICANT: DiFrancesco, Valentina
; APPLICANT: Beasley, Ellen
; TITLE OF INVENTION: ISOLATED HUMAN PROTEASE PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PROTEASE PROTEINS,
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: CL001194
; CURRENT APPLICATION NUMBER: US/09/820,002
; CURRENT FILING DATE: 2001-03-29
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 417

Db 177 FAGAVAQASPHGLQLGQVAVVHHGGLPFRDPNSENNDIALVHLSSPLPLTEYIQPVC 236
QY 309 LPFFDEELTPATPLWILGHFTKQNGKMSDILLOASVQVIDTRCNADDAAYOGVETEM 368
Db 237 LPAAGQALVDGKICTVTGNGT-QYTGQAGVLOEARVPIISNDVCNGADFYGNQIKPM 295
QY 369 MCAGIPEGGVDTCQSGGGLPMYQ-----SDQHVHVGIVSWGCGGPGSTPGVYTKVSAY 423
Db 296 FCAGYPEGGIDACQSGSGGFFVCEDSISRTPRWRLCGIVSWGTCGALAQKPGVYTKVSDF 355
QY 424 LNWYVNVK 432
Db 356 REWIFOAIK 364

RESULT 9

US-09-000-846-2
; Sequence 2, Application US/09000846
; Patent No. 5981830
; GENERAL INFORMATION:
; APPLICANT: WU, QINGYU
; APPLICANT: SADLER, JASPER
; TITLE OF INVENTION: KNOCKOUT MICE AND THEIR PROGENY WITH
; TITLE OF INVENTION: DISRUPTED SERINE PROTEASE GENES
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MILLEN WHITE, ZELANO & BRANIGAN, P.C.
; STREET: 2200 CLARENDON BLVD. SUITE 1400
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: US
; ZIP: 22201
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/000,846
; FILING DATE: 30-DEC-1997
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/866,058
; FILING DATE: 30-MAY-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: LEBOWITZ, RICHARD M.
; REGISTRATION NUMBER: 37,067
; REFERENCE/DOCKET NUMBER: BERLX 65P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-243-6333
; TELEFAX: 703-243-6410
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 416 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-000-846-2

Query Match 24.5%; Score 574; DB 2; Length 416;
Best Local Similarity 30.8%; Pred. No. 3e-50;
Matches 135; Conservative 62; Mismatches 151; Indels 90; Gaps 10;

QY 30 VCIPITIALLSAGIIVVILKVIDKYFLCGOPLHPIRKQLCDGDLCPGLGEDEEH 89
Db 22 VGTLLFLTGIGAASVAIVTILQ-----SDQE- 48
QY 90 CVKSPPEGAVRLSKDRSTLQVLDSATGNWFSACFDNFTALAEATACRMGY----- 143
Db 49 -----PLYQVQSPGDSRLAVFDKTEGTWRLLCSSRSNARVAGLGCENGFRLAH 100
QY 144 -----SSKPTFRAVEIG-----PDQDLVWEITENSQELRMENSSGPCLSGLSVSL 189

Db 101 SELDVRTAGANGTSFFCVDEGLRLAQRLDLVDIVSCD-----CPRGRFLTA 147
QY 190 HCLACG-KSLKTPRVVGGEEASVDSWPQVSIQYDKQHVCGSILDPHWLTAACHCRKH 248
Db 148 TCQDCGRRLKPLVDRIYVGGQDSSLGRWPQVSLRYDGTGLCGGSLLSGDWVLTAAHCFPER 207
QY 249 TDVFN-WKVVRAGSDKLGSPPSLAVAKIIIBFNPMVP-----KNDIALMKLOFPLT 299
Db 208 NRVLGRWRVFAVARTSPHVLQGVQAVIYHGGVLPFRDPTIDENSNDIALVHLSSSLP 267
QY 300 FSGTVRPICLPFFDEELTPATPLWILGHFTKQNGKMSDILLOASVQVIDTRCNADDA 359
Db 268 LLEYIQPVCLPAAGQALVDGKICTVTGNGT-QYTGQAGVLOEARVPIISNEVCNSPDF 326
QY 360 YOGVETEMKAGIPEGGVDTCQSGGGLPMYQ-----SDQHVHVGIVSWGCGGPGSTP 414
Db 327 YGNQIKPMFCAGYPEGGIDACQSGSGGFFVCEDSISRTSRWRLCGIVSWGTCALAKP 386
QY 415 GYVTKVSAYLWYVNVK 432
Db 387 GYVTKVTDFFREWIFRAIK 404

RESULT 10

US-08-200-900A-2
; Sequence 2, Application US/08200900A
; Patent No. 5865586
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: CLONING OF ENTEROKINASE AND METHOD OF USE
; NUMBER OF SEQUENCES: 38
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc. - Legal Affairs
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: MA
; COUNTRY: USA
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/200,900A
; FILING DATE: 23-FEB-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Meinert, Maureen C.
; REGISTRATION NUMBER: 31,544
; REFERENCE/DOCKET NUMBER: GI 5201-FWC
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 876-1170 X8574
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 798 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-200-900A-2

Query Match 24.4%; Score 571; DB 1; Length 798;
Best Local Similarity 35.0%; Pred. No. 1.6e-49;
Matches 133; Conservative 67; Mismatches 138; Indels 42; Gaps 15;
QY 69 IPRKQLCDGDLCPGLGEDEEHCVKSPPEGAVRLSKDRSTLQVLDSATGNWFSACFDN 128
Db 435 IPLVNLCDGPHCFKXGDSAEHCVRLF-NGTTDSSGLVQFR--IQSI-----WHVACAE 485
QY 129 FTEALAEATACRMGY-----SSKPTFRAVEIGPDQDLVVE-----ITENSQELRMENSSGP 180

Db 486 WTQISDDVQCLLGLGTGNSVPTF-STGGGPPYVNLNTAPNGSLILTPSQ----- 535
QY 181 CLSGSLVSLHC--LACGSKLKT-----PRVVGGEASVDSWPMQVSIQYDKQHVCGGSILD 234
Db 536 CLEDSLILLCQNVKSCGKLVQEVSKIVGGSDSRGAPWVVALYFDDQVCGASLVS 595
QY 235 PHWVLTAAHC-FRKHTDVFNNKVRAGSDKLSGF--PSLAVAKIIIEFNPMY---PKDND 288
Db 596 RDLVLSAAHCVYGRNMEPSSKWKAVLGHMASNLTSPOIETRLIDQIVINPHYNKRKKNND 655
QY 289 IALMKLOPFLTFSTGVRPICIPLPFDEBLTTPATPLWIIIGWFTKQNGKMSDILLOASVQV 348
Db 656 IAMHLEMKVNTYDIQIPLPEENOVFPGRICSIAGGALIYQ-GSTADVLOEADVPL 714
QY 349 IDTRCNADDAVQGEVTEKVMCAGIPGGVDTCQDGGGLMYQ-SDQWHVGVIVSWGYG 407
Db 715 LSNEKCO-QQMEYNTENVMVCAGYAGGVDSQCGDGGPLMCQENNRWLLAGVTSFGYQ 773
QY 408 CGGPSTPGVTVKVSAYLNWI 427
Db 774 CALPNRPGVYARVPRFTEWI 793

RESULT 11
PCT-US94-00616-2
; Sequence 2, Application PC/TUS9400616
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: CLONING OF ENTEROKINASE AND METHOD OF USE
; NUMBER OF SEQUENCES: 33
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25 (BPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/00616
; FILING DATE:
; CLASSIFICATION:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 798 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US94-00616-2

Query Match 24.4%; Score 571; DB 5; Length 798;
Best Local Similarity 35.0%; Pred. No. 1.6e-49;
Matches 133; Conservative 67; Mismatches 138; Indels 42; Gaps 15;

QY 69 IPRKQLCDGELDCPLGDEDEHCVKSPFEGPAVAVRLSKDRSTLQVLDSATGNWFSACFDN 128
Db 435 IPLVNLCDGPPHCKDGSDEAHCVRLF-NGTDSGLVQFR--IQSI-----NHVACAEN 485
QY 129 FTEALAEACROMGY-----SSKPTFAVEIGPQDLDVE-----ITNSSELNRNSSGP 180
Db 486 WTQISDDVQCLLGLGTGNSVPTF-STGGGPPYVNLNTAPNGSLILTPSQ----- 535
QY 181 CLSGSLVSLHC--LACGSKLKT-----PRVVGGEASVDSWPMQVSIQYDKQHVCGGSILD 234
Db 536 CLEDSLILLCQNVKSCGKLVQEVSKIVGGSDSRGAPWVVALYFDDQVCGASLVS 595
QY 235 PHWVLTAAHC-FRKHTDVFNNKVRAGSDKLSGF--PSLAVAKIIIEFNPMY---PKDND 288
Db 596 RDLVLSAAHCVYGRNMEPSSKWKAVLGHMASNLTSPOIETRLIDQIVINPHYNKRKKNND 655
QY 289 IALMKLOPFLTFSTGVRPICIPLPFDEBLTTPATPLWIIIGWFTKQNGKMSDILLOASVQV 348
Db 656 IAMHLEMKVNTYDIQIPLPEENOVFPGRICSIAGGALIYQ-GSTADVLOEADVPL 714
QY 349 IDTRCNADDAVQGEVTEKVMCAGIPGGVDTCQDGGGLMYQ-SDQWHVGVIVSWGYG 407

Db 715 LSNEKCO-QQMEYNTENVMVCAGYAGGVDSQCGDGGPLMCQENNRWLLAGVTSFGYQ 773
QY 408 CGGPSTPGVTVKVSAYLNWI 427
Db 774 CALPNRPGVYARVPRFTEWI 793

RESULT 12
US-08-508-448C-25
; Sequence 25, Application US/08508448C
; Patent No. 5804410
; GENERAL INFORMATION:
; APPLICANT: Kazuyoshi YAMAOKA et al.
; TITLE OF INVENTION: NUCLEIC ACID SEQUENCE ENCODING
; TITLE OF INVENTION: TRYPSIN-LIKE ENZYME AND PROCESS FOR PRODUCING THE ENZYME
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack
; STREET: 805 Fifteenth Street, N.W., #700
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/508,448C
; FILING DATE: July 28, 1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Warren M. Cheek, Jr.
; REGISTRATION NUMBER: 33,367
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-8850
; TELEFAX:
; TELEX:
; INFORMATION FOR SEQ ID NO: 25:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 418 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-508-448C-25

Query Match 24.3%; Score 568; DB 1; Length 418;
Best Local Similarity 32.4%; Pred. No. 1.2e-49;
Matches 138; Conservative 79; Mismatches 149; Indels 60; Gaps 14;

QY 33 PIIALLSLASIIIVVLIKVILDKYFICGGLPHIPRK--QLCDGELDCPL----- 83
Db 16 PYWCFIVVAGVILAVTIALLV---YFLAFQKQSVFYRSSFQLLNVEYNSQLNSPATQE 72
QY 84 -----GEDEHCVKSPFEG-----PAVAVRLSKD-----RSTLQVLDSATGNWFSACFD 127
Db 73 YTLTGRIEELITKTFKESNLNQLNFIKRAVAKLRQDGGVRAVVMKFOFTRNNNGASM- 131
QY 128 NTEALAEACROMGYSSKPTFAVEIGPQDLDVEITENSQELMRN--SSGPCLSGS 185
Db 132 ---KRIEIVLQMLNS-----GNLEINP--STETSLTDQAANWLINECAGPDLI-- 180
QY 186 LVSLHCLACGSKLKTFRVVGGEASVDSWPMQVSIQYDKQHVCGGSILDPHWLTAACHF 245
Db 181 -----TLSQRILGTEAEGSWPMQVSLRLNNAHCGGSLINNMLITAAHCF 229
QY 246 RKHTDVFNNKVRAGSDKLSGFPSL--AVAKIIIEFNPMYKNDIALMKLOPFLTFSGT 303

```
Db 230 RSNPNRDMWATSGIST--TFPKLRMRVRLNIHNNYKSATHENDIALVRLNSVTFK 287
QY 304 VRPICLPFFDEELTPATPLMIIGWFTKQNGKMSDILLQASVQVIDSTRCNADDAAYQGE 363
Db 288 IHSVCLPAATQNIIPPGSTAYVTGGAQYAGHTVPE--LRQGVRIISNDVCNAPHSYNGA 346
QY 364 VTEKMCAGIPGEGVDTCQDGGGGLMYQSDQ--WHVVGIVSWGCGGPGSTPGVYTKVS 421
Db 347 ILSGMLCAGVPGQGVADACQDGGGGLVQEDSRRLWFIVIGVSWGDCGLPKFGVYTRV 406
QY 422 AYLNWI 427
Db 407 AYLDWI 412

RESULT 13
US-09-370-838-82
; Sequence 82, Application US/09370838
; Patent No. 6444425
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Mohamath, Roadoh
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOUNDS FOR THERAPY AND DIAGNOSIS OF
; FILE REFERENCE: 210121.475C1
; CURRENT APPLICATION NUMBER: US/09/370,838
; EARLIER FILING DATE: 1999-08-09
; EARLIER APPLICATION NUMBER: US 09/285,323
; NUMBER OF SEQ ID NOS: 289
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 82
; LENGTH: 418
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-370-838-82

Query Match 24.3%; Score 568; DB 4; Length 418;
Best Local Similarity 32.4%; Pred. No. 1.2e-49;
Matches 138; Conservative 79; Mismatches 149; Indels 60; Gaps 14;

QY 33 PIIIALSLASIIIVVLKIVLDKXYFLCGQLHFIPRK--QLCDGELDCPL----- 83
Db 16 PYVWCFIVVAGVILAVTIALIV---YFLAFQKSYFYRSSFQLNVEYNSQLNSPATQ 72
QY 84 -----GEDEHCVKSPFEG-----PAVAVRLSKD-----RSTLQVLDSATGNWFSACFD 127
Db 73 YRTLSGRIESLITKTFKESNLRNQFIRAHVAKLRDQGGVRAVVMKFOFTNNNGASM- 131
QY 128 NFTEALAEATACROMYSSKPTFRAVEIGPDQDLVDVVEITENSQELMRN--SSGPCLSGS 185
Db 132 ---KSRIESVLRQMLNNS---GNLEINP--STEITSITDQAAANWLINECCAGPDLI-- 180
QY 186 LVSLHCLACGSKLTPRYVVGGEASVDSPWQVSIQYDKQHVCGGSILDPHWLTAACHF 245
Db 181 -----TLSEQRILGTEAREEGSPWQVSLRLNNAHHCGLSLNNMWILTAAHCF 229
QY 246 RKHTDVFNVKVRAGSKLGSFPFL--AVAKIIIEFNPMYKNDIALMKLOFFLTFSGT 303
Db 230 RSNPNRDMWATSGIST--TFPKLRMRVRLNIHNNYKSATHENDIALVRLNSVTFK 287
QY 304 VRPICLPFFDEELTPATPLMIIGWFTKQNGKMSDILLQASVQVIDSTRCNADDAAYQGE 363
Db 288 IHSVCLPAATQNIIPPGSTAYVTGGAQYAGHTVPE--LRQGVRIISNDVCNAPHSYNGA 346
QY 364 VTEKMCAGIPGEGVDTCQDGGGGLMYQSDQ--WHVVGIVSWGCGGPGSTPGVYTKVS 421
Db 347 ILSGMLCAGVPGQGVADACQDGGGGLVQEDSRRLWFIVIGVSWGDCGLPKFGVYTRV 406
QY 422 AYLNWI 427
Db 407 AYLDWI 412

RESULT 15
US-09-370-838-62
; Sequence 62, Application US/09370838
; Patent No. 6444425
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Mohamath, Roadoh
```

```
Db 407 AYLDWI 412

RESULT 14
US-09-370-838-83
; Sequence 83, Application US/09370838
; Patent No. 6444425
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Mohamath, Roadoh
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOUNDS FOR THERAPY AND DIAGNOSIS OF
; FILE REFERENCE: 210121.475C1
; CURRENT APPLICATION NUMBER: US/09/370,838
; EARLIER FILING DATE: 1999-08-09
; EARLIER APPLICATION NUMBER: US 09/285,323
; NUMBER OF SEQ ID NOS: 289
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 83
; LENGTH: 418
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-370-838-83

Query Match 24.3%; Score 568; DB 4; Length 418;
Best Local Similarity 32.4%; Pred. No. 1.2e-49;
Matches 138; Conservative 79; Mismatches 149; Indels 60; Gaps 14;

QY 33 PIIIALSLASIIIVVLKIVLDKXYFLCGQLHFIPRK--QLCDGELDCPL----- 83
Db 16 PYVWCFIVVAGVILAVTIALIV---YFLAFQKSYFYRSSFQLNVEYNSQLNSPATQ 72
QY 84 -----GEDEHCVKSPFEG-----PAVAVRLSKD-----RSTLQVLDSATGNWFSACFD 127
Db 73 YRTLSGRIESLITKTFKESNLRNQFIRAHVAKLRDQGGVRAVVMKFOFTNNNGASM- 131
QY 128 NFTEALAEATACROMYSSKPTFRAVEIGPDQDLVDVVEITENSQELMRN--SSGPCLSGS 185
Db 132 ---KSRIESVLRQMLNNS---GNLEINP--STEITSITDQAAANWLINECCAGPDLI-- 180
QY 186 LVSLHCLACGSKLTPRYVVGGEASVDSPWQVSIQYDKQHVCGGSILDPHWLTAACHF 245
Db 181 -----TLSEQRILGTEAREEGSPWQVSLRLNNAHHCGLSLNNMWILTAAHCF 229
QY 246 RKHTDVFNVKVRAGSKLGSFPFL--AVAKIIIEFNPMYKNDIALMKLOFFLTFSGT 303
Db 230 RSNPNRDMWATSGIST--TFPKLRMRVRLNIHNNYKSATHENDIALVRLNSVTFK 287
QY 304 VRPICLPFFDEELTPATPLMIIGWFTKQNGKMSDILLQASVQVIDSTRCNADDAAYQGE 363
Db 288 IHSVCLPAATQNIIPPGSTAYVTGGAQYAGHTVPE--LRQGVRIISNDVCNAPHSYNGA 346
QY 364 VTEKMCAGIPGEGVDTCQDGGGGLMYQSDQ--WHVVGIVSWGCGGPGSTPGVYTKVS 421
Db 347 ILSGMLCAGVPGQGVADACQDGGGGLVQEDSRRLWFIVIGVSWGDCGLPKFGVYTRV 406
QY 422 AYLNWI 427
Db 407 AYLDWI 412

RESULT 15
US-09-370-838-62
; Sequence 62, Application US/09370838
; Patent No. 6444425
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Mohamath, Roadoh
```



```
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOUNDS FOR THERAPY AND DIAGNOSIS OF
; TITLE OF INVENTION: LUNG CANCER AND METHODS FOR THEIR USE
; FILE REFERENCE: 210121.475C1
; CURRENT APPLICATION NUMBER: US/09/370.838
; CURRENT FILING DATE: 1999-08-09
; EARLIER APPLICATION NUMBER: US 09/285.323
; EARLIER FILING DATE: 1999-04-02
; NUMBER OF SEQ ID NOS: 289
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 62
; LENGTH: 418
; TYPE: PRT
; ORGANISM: Homo sapien
; US-09-370-838-62

Query Match      24.2%; Score 566; DB 4; Length 418;
Best Local Similarity 32.2%; Pred. No. 2e-49;
Matches 137; Conservative 80; Mismatches 149; Indels 60; Gaps 14;

QY 33 PIIIALSLASIIIVVVLKIVLDKYFLCGQPLHFIPRK--QLCDGELDCPL----- 83
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 16 PYVVCFTWAGVILAVTIALV--YFLAFDQKSYFYRSSFOLLNVEYNSQLNSPATQE 72
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 84 -----GEDEEHCVKSPFEG-----PAVAVRLSKD-----RSTLOVLDSATGNWFSACPD 127
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 73 YRLTSGRIEISLITKTFKESNLNRFIRAHVAKLQDGSVGRADYVMKFQFTRNNNGASM- 131
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 128 NFEALAAETACROMGYSSKPTFRAVEIGTQDQDLVDWEITENSQELWRN--SSGPCLSGS 185
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 132 ---KSRIESVLRQMLNNS-----GNLEINP--STEITSLTQAAANWLINCEGAGPDLI-- 180
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 186 LVSLHCLACGSKLTPRVVGGEERASVDSPWQVSIQYDKOHVCGGSILDPHWVLTAAHCF 245
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 181 -----TLSEQRILGTEAEGSPWQVSLRLNNAHHCGLSLINNMMWILTAAHCF 229
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 246 RKHTDVFNWVRAGSDKLGSFPSSL--AVAKIIILEFNPYPKNDIALMKLPPLTFESGT 303
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 230 RSNPNRPDWIATSGIST--TFPKLRMVRNLIHNNYKSAETHENDIALVRLNSVTFTKD 287
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 304 VRPICLPFFDEELTPATPLMIIGWFTKQNGKMSDILLQASVQVIDSTRCNADDDAYQGE 363
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 288 IHSVCLPAATQNIPTPGSTAYVTGWGAQYAGHTYPE-LRQGVRIISNDVCNAPHSYNGA 346
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 364 VTERMCAGIPEGVDPCQDSGGLMYQSDQ--WHVVGVSMGYCGGSPSTPGVYTKVS 421
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 347 ILSGMLCAGVPGGVDACQDSGGLVQEDSRRLWFIVGVSWGDQCGLEPKPGVYTRVT 406
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 422 AYLNWI 427
   : : : : :
Db 407 AYIDWI 412
   : : : : :
```

Search completed: November 23, 2003, 07:53:46
Job time : 25.1389 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 23, 2003, 07:49:15 ; Search time 27.5241 Seconds
(without alignments)
2885.234 Million cell updates/sec

Title: US-09-607-745-2

Perfect score: 2342

Sequence: 1 MDPDSQPLNSLDVKPLRKP.....VTKVSAYLNWYVWKAL 435

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 666188 seqs, 182559486 residues

Total number of hits satisfying chosen parameters: 666188

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW PUB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US09C_NEW PUB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubpaa/US10C_NEW PUB.pep.*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW PUB.pep.*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	ID	Description
1	2338	99.8	435	12	US-10-180-719-6
2	2337	99.8	437	9	US-09-851-588-8
3	2319	99.0	437	11	US-09-776-191-4
4	2297.5	98.1	432	11	US-09-888-257A-7
5	2297.5	98.1	432	11	US-09-946-374-275
6	2297.5	98.1	432	12	US-10-015-387A-275
7	2297.5	98.1	432	12	US-10-063-735-112
8	2297.5	98.1	432	12	US-10-006-130A-275
9	2297.5	98.1	432	12	US-10-199-672-330
10	2297.5	98.1	432	12	US-10-006-172A-275
11	2297.5	98.1	432	12	US-10-187-749-330
12	2297.5	98.1	432	12	US-10-194-457-330
13	2297.5	98.1	432	12	US-10-184-642-330
14	2297.5	98.1	432	12	US-10-196-747-330
15	2297.5	98.1	432	12	US-10-015-392A-275

16	2297.5	98.1	432	12	US-10-017-253A-275
17	2297.5	98.1	432	12	US-10-173-689-330
18	2297.5	98.1	432	12	US-10-173-690-330
19	2297.5	98.1	432	12	US-10-173-691-330
20	2297.5	98.1	432	12	US-10-173-692-330
21	2297.5	98.1	432	12	US-10-173-693-330
22	2297.5	98.1	432	12	US-10-173-698-330
23	2297.5	98.1	432	12	US-10-173-699-330
24	2297.5	98.1	432	12	US-10-173-707-330
25	2297.5	98.1	432	12	US-10-174-569-330
26	2297.5	98.1	432	12	US-10-174-583-330
27	2297.5	98.1	432	12	US-10-174-587-330
28	2297.5	98.1	432	12	US-10-174-589-330
29	2297.5	98.1	432	12	US-10-174-591-330
30	2297.5	98.1	432	12	US-10-175-736-330
31	2297.5	98.1	432	12	US-10-175-742-330
32	2297.5	98.1	432	12	US-10-175-744-330
33	2297.5	98.1	432	12	US-10-175-745-330
34	2297.5	98.1	432	12	US-10-175-746-330
35	2297.5	98.1	432	12	US-10-175-751-330
36	2297.5	98.1	432	12	US-10-175-754-330
37	2297.5	98.1	432	12	US-10-176-480-330
38	2297.5	98.1	432	12	US-10-176-489-330
39	2297.5	98.1	432	12	US-10-176-754-330
40	2297.5	98.1	432	12	US-10-176-755-330
41	2297.5	98.1	432	12	US-10-176-759-330
42	2297.5	98.1	432	12	US-10-176-920-330
43	2297.5	98.1	432	12	US-10-176-922-330
44	2297.5	98.1	432	12	US-10-176-924-330
45	2297.5	98.1	432	12	US-10-176-984-330

ALIGNMENTS

RESULT 1

US-10-180-719-6

; Sequence 6, Application US/10180719

; Publication No. US20030166246A1

; GENERAL INFORMATION:

; APPLICANT: Bandman, Olga

; Hillman, Jennifer L.

; Yue, Henry

; Guegler, Karl J.

; Corley, Neil C.

; Tang, Tom Y.

; Shah, Purvi

; TITLE OF INVENTION: HUMAN PROTEASE MOLECULES

; NUMBER OF SEQUENCES: 24

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Incyte Pharmaceuticals, Inc.

; STREET: 3174 Porter Dr.

; CITY: Palo Alto

; STATE: CA

; COUNTRY: USA

; ZIP: 94304

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; SOFTWARE: FastSeq for Windows Version 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/10/180,719

; FILING DATE: 25-Jun-2002

; PRIORITY INFORMATION DATA:

; APPLICATION NUMBER: US/09/008,271

; FILING DATE: 16-Jan-1998

; ATTORNEY/AGENT INFORMATION:

; NAME: Mohan-Peterson, Sheela

; REGISTRATION NUMBER: 41,201

; REFERENCE/DOCKET NUMBER: PF-0458 US

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 650-855-0555

TELEFAX: 650-845-4166

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:

LENGTH: 435 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

IMMEDIATE SOURCE:

LIBRARY: COLNONT13

CLONE: 1337018

SEQUENCE DESCRIPTION: SEQ ID NO: 6 :

US-10-180-719-6

Query Match 99.8%; Score 2338; DB 12; Length 435;
Best Local Similarity 99.8%; Pred. No. 5.2e-226;
Matches 434; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDPDSQPLNSLDVVKPLRKPRIPMETFRKVGIIPIIIALLSLASIIIVVLIKVILDKYIF 60
DB 1 MDPDSQPLNSLDVVKPLRKPRIPMETFRKVGIIPIIIALLSLASIIIVVLIKVILDKYIF 60

QY 61 LCQPLHFIPRKQKQELDCPLGDEDEHCVKFPPEGPAVAVRLSKDRSTLOVLDGATGN 120
DB 61 LCQPLHFIPRKQKQELDCPLGDEDEHCVKFPPEGPAVAVRLSKDRSTLOVLDGATGN 120

QY 121 WFSACPDNTEALAEACRQMGYSKPTTFAVIGDPDQDLDVVEITENSQELMRNSSGP 180
DB 121 WFSACPDNTEALAEACRQMGYSKPTTFAVIGDPDQDLDVVEITENSQELMRNSSGP 180

QY 181 CLSGSLVSLHCLACGSLKTPRVVGGEEASVDSWPQVSIQYDKQHVCGGSIIDPHWLT 240
DB 181 CLSGSLVSLHCLACGSLKTPRVVGGEEASVDSWPQVSIQYDKQHVCGGSIIDPHWLT 240

QY 241 AAHCFRKHTDVFNNKVRAGSDKLGSPFLAVAKIIIEFNPMYPKNDIALMKLOPLTF 300
DB 241 AAHCFRKHTDVFNNKVRAGSDKLGSPFLAVAKIIIEFNPMYPKNDIALMKLOPLTF 300

QY 301 SGTVRPICLPFFDEBELTPATPLMIIGWFTKONGKMSDILLQASVQVIDSTRCNADDA 360
DB 301 SGTVRPICLPFFDEBELTPATPLMIIGWFTKONGKMSDILLQASVQVIDSTRCNADDA 360

QY 361 QGEVTEKMCAGIPGEGVDTCCQDSGGPLMYQSDQHVHVGIVSWGVCGGSPSTPGVYTKV 420
DB 361 QGEVTEKMCAGIPGEGVDTCCQDSGGPLMYQSDQHVHVGIVSWGVCGGSPSTPGVYTKV 420

QY 421 SAYLNWIYNWKAEL 435
DB 421 SAYLNWIYNWKAEL 435

RESULT 2

US-09-851-588-8
; Sequence 8, Application US/09851588
; Patent No. US20020042067A1
; GENERAL INFORMATION:
; APPLICANT: Mack, David
; APPLICANT: Gish, Kurt C.
; APPLICANT: Wilson, Keith E.
; TITLE OF INVENTION: NOVEL METHODS OF DIAGNOSING COLORECTAL CANCER, COMPOSITIONS, AND
; FILE OF INVENTION: OF SCREENING FOR COLORECTAL CANCER MODULATORS
; FILE REFERENCE: A-68829-1/DUB/JJD/AMS
; CURRENT APPLICATION NUMBER: US/09/851,588
; CURRENT FILING DATE: 2001-09-24
; PRIOR APPLICATION NUMBER: US 09/642,252
; PRIOR FILING DATE: 2000-08-17
; PRIOR APPLICATION NUMBER: US 09/656,002
; PRIOR FILING DATE: 2000-09-06
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8
; LENGTH: 437
; TYPE: PRT
; ORGANISM: Homo sapiens

US-09-851-588-8

Query Match 99.8%; Score 2337; DB 9; Length 437;
Best Local Similarity 100.0%; Pred. No. 6.6e-226;
Matches 434; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 DPDSQPLNSLDVVKPLRKPRIPMETFRKVGIIPIIIALLSLASIIIVVLIKVILDKYIF 61
DB 4 DPDSQPLNSLDVVKPLRKPRIPMETFRKVGIIPIIIALLSLASIIIVVLIKVILDKYIF 63

QY 62 CGOPLHFIPRKQKQELDCPLGDEDEHCVKFPPEGPAVAVRLSKDRSTLOVLDGATGN 121
DB 64 CGOPLHFIPRKQKQELDCPLGDEDEHCVKFPPEGPAVAVRLSKDRSTLOVLDGATGN 123

QY 122 FSACPDNTEALAEACRQMGYSKPTTFAVIGDPDQDLDVVEITENSQELMRNSSGP 181
DB 124 FSACPDNTEALAEACRQMGYSKPTTFAVIGDPDQDLDVVEITENSQELMRNSSGP 183

QY 182 LSGSLVSLHCLACGSLKTPRVVGGEEASVDSWPQVSIQYDKQHVCGGSIIDPHWLT 241
DB 184 LSGSLVSLHCLACGSLKTPRVVGGEEASVDSWPQVSIQYDKQHVCGGSIIDPHWLT 243

QY 242 AAHCFRKHTDVFNNKVRAGSDKLGSPFLAVAKIIIEFNPMYPKNDIALMKLOPLTF 301
DB 244 AAHCFRKHTDVFNNKVRAGSDKLGSPFLAVAKIIIEFNPMYPKNDIALMKLOPLTF 303

QY 302 GTVRPICLPFFDEBELTPATPLMIIGWFTKONGKMSDILLQASVQVIDSTRCNADDA 361
DB 304 GTVRPICLPFFDEBELTPATPLMIIGWFTKONGKMSDILLQASVQVIDSTRCNADDA 363

QY 362 GVTETKMCAGIPGEGVDTCCQDSGGPLMYQSDQHVHVGIVSWGVCGGSPSTPGVYTKV 421
DB 364 GVTETKMCAGIPGEGVDTCCQDSGGPLMYQSDQHVHVGIVSWGVCGGSPSTPGVYTKV 423

QY 422 AYLNWIYNWKAEL 435
DB 424 AYLNWIYNWKAEL 437

RESULT 3

US-09-776-191-4
; Sequence 4, Application US/09776191
; Publication No. US20030119168A1
; GENERAL INFORMATION:
; APPLICANT: Edwin L. Madison
; APPLICANT: Edgar O. Ong
; APPLICANT: Jiunn-Chern Yeh
; APPLICANT: Corvas International, Inc.
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING
; FILE OF INVENTION: TRANSMEMBRANE SERINE PROTEASES, THE ENCODED PROTEINS AND
; FILE REFERENCE: 24745-1607
; CURRENT APPLICATION NUMBER: US/09/776,191
; CURRENT FILING DATE: 2001-02-02
; PRIOR APPLICATION NUMBER: 60/213,124
; PRIOR FILING DATE: 2000-06-22
; PRIOR APPLICATION NUMBER: 60/234,840
; PRIOR FILING DATE: 2000-06-22
; PRIOR APPLICATION NUMBER: 60/179,982
; PRIOR FILING DATE: 2000-02-03
; PRIOR APPLICATION NUMBER: 60/183,542
; PRIOR FILING DATE: 2000-02-18
; PRIOR APPLICATION NUMBER: 09/657,968
; PRIOR FILING DATE: 2000-02-08
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 437
; TYPE: PRT
; ORGANISM: Homo Sapien

US-09-776-191-4

Query Match 99.0%; Score 2319; DB 11; Length 437;

Best Local Similarity 99.5%; Pred. No. 4.3e-224;
Matches 432; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 2 DPDSQPLNSLDVVKPLRKPRIPMETFRKVGIPPIIIALLSLASIIIVVLIKVLKYYFL 61
Ddb 4 DPDSQPLNSLDVVKPLRKPRIPMETFRKVGIPPIIIALLSLASIIIVVLIKVLKYYFL 63
QY 62 CQOPLHFIPRKQLCDGELDCPLGEDDEHCVKSPPEGPAVAVRLSKDRSTLOVLDSATGNW 121
Ddb 64 CQOPLHFIPRKQLCDGELDCPLGEDDEHCVKSPPEGPAVAVRLSKDRSTLOVLDSATGNW 123
QY 122 FSACFDNTEALAEACRQMGYSKPTFRAVEIGPDQDLVVEITENSQELMRNSSGFC 181
Ddb 124 FSACFDNTEALAEACRQMGYSKPTFRAVEIGPDQDLVVEITENSQELMRNSSGFC 183
QY 182 LSGSLVSLHCLACGKSLKTRPVVVGEEASVDSWPMQVSIQYDKQHVCGGSIILDPHWLTA 241
Ddb 184 LSGSLVSLHCLACGKSLKTRPVVVGEEASVDSWPMQVSIQYDKQHVCGGSIILDPHWLTA 243
QY 242 AHCERKHTDVFNMKVRAGSKLGSFSLAVAKIIIEFNPMYPKNDIALMKLOPFLTES 301
Ddb 244 AHCERKHTDVFNMKVRAGSKLGSFSLAVAKIIIEFNPMYPKNDIALMKLOPFLTES 303
QY 302 GTVRPICLPFFDEELTPATPLWIIQWGTCKONGKMSDILLQASVOVIDSTRCNADDAQ 361
Ddb 304 GTVRPICLPFFDEELTPATPLWIIQWGTCKONGKMSDILLQASVOVIDSTRCNADDAQ 363
QY 362 GEVTEKMCAGIPEGGVDTCCQDGGGFLMYQSDQHVGVISWGYCGGSPSTPGVYTKVS 421
Ddb 364 GEVTEKMCAGIPEGGVDTCCQDGGGFLMYQSDQHVGVISWGYCGGSPSTPGVYTKVS 423
QY 422 AYLNWIYNNVWKAEL 435
Ddb 424 AYLNWIYNNVWKAEL 437

RESULT 4

US-09-888-257A-7

; Sequence 7, Application US/09888257A

; Publication No. US2003060612A1

; GENERAL INFORMATION:

; APPLICANT: Goddard, Audrey

; APPLICANT: Godowski, Paul J.

; APPLICANT: Gurney, Austin L.

; APPLICANT: Hillan, Kenneth J.

; APPLICANT: Polakis, Paul

; APPLICANT: Smith, Victoria

; APPLICANT: Wood, William I.

; APPLICANT: Wu, Thomas D.

; APPLICANT: Zhang, Zemin

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DIAGNOSIS AND

; TITILE OF INVENTION: TREATMENT OF TUMOR

; FILE REFERENCE: P5002R1

; CURRENT APPLICATION NUMBER: US/09/888,257A

; CURRENT FILING DATE: 2002-04-30

; PRIOR APPLICATION NUMBER: US 60/063,540

; PRIOR FILING DATE: 1997-10-28

; PRIOR APPLICATION NUMBER: US 60/089,553

; PRIOR FILING DATE: 1998-06-17

; PRIOR APPLICATION NUMBER: US 60/099,792

; PRIOR FILING DATE: 1998-09-10

; PRIOR APPLICATION NUMBER: US 60/103,678

; PRIOR FILING DATE: 1998-10-08

; PRIOR APPLICATION NUMBER: US 60/235,451

; PRIOR FILING DATE: 2000-09-26

; PRIOR APPLICATION NUMBER: PCT/US99/12252

; PRIOR FILING DATE: 1999-06-02

; PRIOR APPLICATION NUMBER: PCT/US99/20111

; PRIOR FILING DATE: 1999-09-01

; PRIOR APPLICATION NUMBER: PCT/US00/04342

; PRIOR FILING DATE: 2000-02-18

; PRIOR APPLICATION NUMBER: PCT/US00/05841

; PRIOR FILING DATE: 2000-03-02

; PRIOR APPLICATION NUMBER: PCT/US00/08439
; PRIOR FILING DATE: 2000-03-30
; PRIOR APPLICATION NUMBER: PCT/US00/23328
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: PCT/US00/32678
; PRIOR FILING DATE: 2000-12-01
; PRIOR APPLICATION NUMBER: PCT/US01/06520
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: PCT/US01/06666
; PRIOR FILING DATE: 2001-03-01
; NUMBER OF SEQ ID NOS: 10
; SEQ ID NO 7
; LENGTH: 432
; TYPE: PRT
; ORGANISM: Homo Sapien
; US-09-888-257A-7
Query Match 98.1%; Score 2297.5; DB 11; Length 432;
Best Local Similarity 98.8%; Pred. No. 6.1e-222;
Matches 429; Conservative 0; Mismatches 0; Indels 5; Gaps 1;
QY 2 DPDSQPLNSLDVVKPLRKPRIPMETFRKVGIPPIIIALLSLASIIIVVLIKVLKYYFL 61
Ddb 4 DPDSQPLNSLDVVKPLRKPRIPMETFRKVGIPPIIIALLSLASIIIVVLIKVLKYYFL 63
QY 62 CQOPLHFIPRKQLCDGELDCPLGEDDEHCVKSPPEGPAVAVRLSKDRSTLOVLDSATGNW 121
Ddb 64 CQOPLHFIPRKQLCDGELDCPLGEDDEHCVKSPPEGPAVAVRLSKDRSTLOVLDSATGNW 123
QY 122 FSACFDNTEALAEACRQMGYSKPTFRAVEIGPDQDLVVEITENSQELMRNSSGFC 181
Ddb 124 FSACFDNTEALAEACRQMGYSKPTFRAVEIGPDQDLVVEITENSQELMRNSSGFC 178
QY 182 LSGSLVSLHCLACGKSLKTRPVVVGEEASVDSWPMQVSIQYDKQHVCGGSIILDPHWLTA 241
Ddb 179 LSGSLVSLHCLACGKSLKTRPVVVGEEASVDSWPMQVSIQYDKQHVCGGSIILDPHWLTA 238
QY 242 AHCERKHTDVFNMKVRAGSKLGSFSLAVAKIIIEFNPMYPKNDIALMKLOPFLTES 301
Ddb 239 AHCERKHTDVFNMKVRAGSKLGSFSLAVAKIIIEFNPMYPKNDIALMKLOPFLTES 298
QY 302 GTVRPICLPFFDEELTPATPLWIIQWGTCKONGKMSDILLQASVOVIDSTRCNADDAQ 361
Ddb 299 GTVRPICLPFFDEELTPATPLWIIQWGTCKONGKMSDILLQASVOVIDSTRCNADDAQ 358
QY 362 GEVTEKMCAGIPEGGVDTCCQDGGGFLMYQSDQHVGVISWGYCGGSPSTPGVYTKVS 421
Ddb 359 GEVTEKMCAGIPEGGVDTCCQDGGGFLMYQSDQHVGVISWGYCGGSPSTPGVYTKVS 418
QY 422 AYLNWIYNNVWKAEL 435
Ddb 419 AYLNWIYNNVWKAEL 432

RESULT 5

US-09-946-374-275

; Sequence 275, Application US/09946374

; Publication No. US20030073129A1

; GENERAL INFORMATION:

; APPLICANT: Baker, Kevin P.

; APPLICANT: Botstein, David

; APPLICANT: Desnoyers, Luc

; APPLICANT: Eaton, Dan L.

; APPLICANT: Ferrara, Napoleone

; APPLICANT: Fong, Sherman

; APPLICANT: Gao, Wei-Qiang

; APPLICANT: Goddard, Audrey

; APPLICANT: Godowski, Paul J.

; APPLICANT: Grimaldi, Christopher J.

; APPLICANT: Gurney, Austin L.

; APPLICANT: Hillan, Kenneth J.

; APPLICANT: Pan, James

; APPLICANT: Paoni, Nicholas F.

APPLICANT: Roy, Margaret Ann
APPLICANT: Smith, Victoria
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Watanabe, Colin K.
APPLICANT: Williams, P. Mickey
APPLICANT: Wood, William I.
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
FILE REFERENCE: P2830P1C1
CURRENT APPLICATION NUMBER: US/09/946,374
CURRENT FILING DATE: 2001-09-04
PRIOR APPLICATION NUMBER: 60/098716
PRIOR FILING DATE: 1998-09-01
PRIOR APPLICATION NUMBER: 60/098723
PRIOR FILING DATE: 1998-09-01
PRIOR APPLICATION NUMBER: 60/098749
PRIOR FILING DATE: 1998-09-01
PRIOR APPLICATION NUMBER: 60/098750
PRIOR FILING DATE: 1998-09-01
PRIOR APPLICATION NUMBER: 60/098803
PRIOR FILING DATE: 1998-09-02
PRIOR APPLICATION NUMBER: 60/098821
PRIOR FILING DATE: 1998-09-02
PRIOR APPLICATION NUMBER: 60/098843
PRIOR FILING DATE: 1998-09-02
PRIOR APPLICATION NUMBER: 60/099536
PRIOR FILING DATE: 1998-09-09
PRIOR APPLICATION NUMBER: 60/099596
PRIOR FILING DATE: 1998-09-09
PRIOR APPLICATION NUMBER: 60/099598
PRIOR FILING DATE: 1998-09-09
PRIOR APPLICATION NUMBER: 60/099602
PRIOR FILING DATE: 1998-09-09
PRIOR APPLICATION NUMBER: 60/099642
PRIOR FILING DATE: 1998-09-09
PRIOR APPLICATION NUMBER: 60/099741
PRIOR FILING DATE: 1998-09-10
PRIOR APPLICATION NUMBER: 60/099754
PRIOR FILING DATE: 1998-09-10
PRIOR APPLICATION NUMBER: 60/099763
PRIOR FILING DATE: 1998-09-10
PRIOR APPLICATION NUMBER: 60/099792
PRIOR FILING DATE: 1998-09-10
PRIOR APPLICATION NUMBER: 60/099808
PRIOR FILING DATE: 1998-09-10
PRIOR APPLICATION NUMBER: 60/099812
PRIOR FILING DATE: 1998-09-10
PRIOR APPLICATION NUMBER: 60/099815
PRIOR FILING DATE: 1998-09-10
PRIOR APPLICATION NUMBER: 60/099816
PRIOR FILING DATE: 1998-09-10
PRIOR APPLICATION NUMBER: 60/100385
PRIOR FILING DATE: 1998-09-15
PRIOR APPLICATION NUMBER: 60/100388
PRIOR FILING DATE: 1998-09-15
PRIOR APPLICATION NUMBER: 60/100390
PRIOR FILING DATE: 1998-09-15
PRIOR APPLICATION NUMBER: 60/100584
PRIOR FILING DATE: 1998-09-16
PRIOR APPLICATION NUMBER: 60/100627
PRIOR FILING DATE: 1998-09-16
PRIOR APPLICATION NUMBER: 60/100661
PRIOR FILING DATE: 1998-09-16
PRIOR APPLICATION NUMBER: 60/100662
PRIOR FILING DATE: 1998-09-16
PRIOR APPLICATION NUMBER: 60/100664
PRIOR FILING DATE: 1998-09-16
PRIOR APPLICATION NUMBER: 60/100683
PRIOR FILING DATE: 1998-09-17
PRIOR APPLICATION NUMBER: 60/100684
PRIOR FILING DATE: 1998-09-17
PRIOR APPLICATION NUMBER: 60/100710
PRIOR FILING DATE: 1998-09-17
PRIOR APPLICATION NUMBER: 60/100711
PRIOR FILING DATE: 1998-09-17
PRIOR APPLICATION NUMBER: 60/100848
PRIOR FILING DATE: 1998-09-18
PRIOR APPLICATION NUMBER: 60/100849
PRIOR FILING DATE: 1998-09-18
PRIOR APPLICATION NUMBER: 60/100919
PRIOR FILING DATE: 1998-09-17
PRIOR APPLICATION NUMBER: 60/100930
PRIOR FILING DATE: 1998-09-17
PRIOR APPLICATION NUMBER: 60/101014
PRIOR FILING DATE: 1998-09-18
PRIOR APPLICATION NUMBER: 60/101068
PRIOR FILING DATE: 1998-09-18
PRIOR APPLICATION NUMBER: 60/101071
PRIOR FILING DATE: 1998-09-18
PRIOR APPLICATION NUMBER: 60/101279
PRIOR FILING DATE: 1998-09-22
PRIOR APPLICATION NUMBER: 60/101471
PRIOR FILING DATE: 1998-09-23
PRIOR APPLICATION NUMBER: 60/101472
PRIOR FILING DATE: 1998-09-23
PRIOR APPLICATION NUMBER: 60/101474
PRIOR FILING DATE: 1998-09-23
PRIOR APPLICATION NUMBER: 60/101475
PRIOR FILING DATE: 1998-09-23
PRIOR APPLICATION NUMBER: 60/101476
PRIOR FILING DATE: 1998-09-23
PRIOR APPLICATION NUMBER: 60/101477
PRIOR FILING DATE: 1998-09-23
PRIOR APPLICATION NUMBER: 60/101479
PRIOR FILING DATE: 1998-09-23
PRIOR APPLICATION NUMBER: 60/101738
PRIOR FILING DATE: 1998-09-24
PRIOR APPLICATION NUMBER: 60/101741
PRIOR FILING DATE: 1998-09-24
PRIOR APPLICATION NUMBER: 60/101743
PRIOR FILING DATE: 1998-09-24
PRIOR APPLICATION NUMBER: 60/101915
PRIOR FILING DATE: 1998-09-24
PRIOR APPLICATION NUMBER: 60/101916
PRIOR FILING DATE: 1998-09-24
PRIOR APPLICATION NUMBER: 60/102207
PRIOR FILING DATE: 1998-09-29
PRIOR APPLICATION NUMBER: 60/102240
PRIOR FILING DATE: 1998-09-29
PRIOR APPLICATION NUMBER: 60/102307
PRIOR FILING DATE: 1998-09-29
PRIOR APPLICATION NUMBER: 60/102330
PRIOR FILING DATE: 1998-09-29
PRIOR APPLICATION NUMBER: 60/102331
PRIOR FILING DATE: 1998-09-29
PRIOR APPLICATION NUMBER: 60/102484
PRIOR FILING DATE: 1998-09-30
PRIOR APPLICATION NUMBER: 60/102487
PRIOR FILING DATE: 1998-09-30
PRIOR APPLICATION NUMBER: 60/102570
PRIOR FILING DATE: 1998-09-30
PRIOR APPLICATION NUMBER: 60/102571
PRIOR FILING DATE: 1998-09-30
PRIOR APPLICATION NUMBER: 60/102684
PRIOR FILING DATE: 1998-10-01
PRIOR APPLICATION NUMBER: 60/102687
PRIOR FILING DATE: 1998-10-01
PRIOR APPLICATION NUMBER: 60/102965
PRIOR FILING DATE: 1998-10-02
PRIOR APPLICATION NUMBER: 60/103258
PRIOR FILING DATE: 1998-10-06
PRIOR APPLICATION NUMBER: 60/103314
PRIOR FILING DATE: 1998-10-07
PRIOR APPLICATION NUMBER: 60/103315
PRIOR FILING DATE: 1998-10-07

PRIOR APPLICATION NUMBER: 60/103328
PRIOR FILING DATE: 1998-10-07
PRIOR APPLICATION NUMBER: 60/103395
PRIOR FILING DATE: 1998-10-07
PRIOR APPLICATION NUMBER: 60/103396
PRIOR FILING DATE: 1998-10-07
PRIOR APPLICATION NUMBER: 60/103401
PRIOR FILING DATE: 1998-10-07
PRIOR APPLICATION NUMBER: 60/103449
PRIOR FILING DATE: 1998-10-06
PRIOR APPLICATION NUMBER: 60/103633
PRIOR FILING DATE: 1998-10-08
PRIOR APPLICATION NUMBER: 60/103678
PRIOR FILING DATE: 1998-10-08
PRIOR APPLICATION NUMBER: 60/103679
PRIOR FILING DATE: 1998-10-08
PRIOR APPLICATION NUMBER: 60/103711
PRIOR FILING DATE: 1998-10-08
PRIOR APPLICATION NUMBER: 60/104257
PRIOR FILING DATE: 1998-10-14
PRIOR APPLICATION NUMBER: 60/104987
PRIOR FILING DATE: 1998-10-20
PRIOR APPLICATION NUMBER: 60/105000
PRIOR FILING DATE: 1998-10-20
PRIOR APPLICATION NUMBER: 60/105002
PRIOR FILING DATE: 1998-10-20
PRIOR APPLICATION NUMBER: 60/105104
PRIOR FILING DATE: 1998-10-21
PRIOR APPLICATION NUMBER: 60/105169
PRIOR FILING DATE: 1998-10-22
PRIOR APPLICATION NUMBER: 60/105266
PRIOR FILING DATE: 1998-10-22
PRIOR APPLICATION NUMBER: 60/105693
PRIOR FILING DATE: 1998-10-26
PRIOR APPLICATION NUMBER: 60/105694
PRIOR FILING DATE: 1998-10-26
PRIOR APPLICATION NUMBER: 60/105807

Query Match 98.1%; Score 2297.5; DB 11; Length 432;
Best Local Similarity 98.8%; Pred. No. 6.1e-222;
Matches 429; Conservative 0; Mismatches 0; Indels 5; Gaps 1;
QY 2 DPDSQPLNSLDVKPRIPMETFRKVGIPPIIIALLSLASIIIVVLIKVLIDKYFL 61
DB 4 DPDSQPLNSLDVKPRIPMETFRKVGIPPIIIALLSLASIIIVVLIKVLIDKYFL 63
QY 62 CGQPLHFIPRKQDCGELDCPLGEDEBEHCVKSPPEGPAVAVRLSKDRSTLQVLDSATGNW 121
DB 64 CGQPLHFIPRKQDCGELDCPLGEDEBEHCVKSPPEGPAVAVRLSKDRSTLQVLDSATGNW 123
QY 122 FSACFDNFTALAEACRQMGYSKPTTFAVEIGPDQDLDVVEITENSQELMRNSSGFC 181
DB 124 FSACFDNFTALAEACRQMGYSKPTTFAVEIGPDQDLDVVEITENSQELMRNSSGFC 178
QY 182 LSGSLVSLHCLACGKSLKTPRVVGGEEASVDSWPQVSIQYDKQHVCGGSIIDPHWVLT 241
DB 179 LSGSLVSLHCLACGKSLKTPRVVGGEEASVDSWPQVSIQYDKQHVCGGSIIDPHWVLT 238
QY 242 AHCPRKHTDVFNNKVRAGSKLGSPFSLAVAKIIIEFNPMYPKNDIALMKLOPPLTFS 301
DB 239 AHCPRKHTDVFNNKVRAGSKLGSPFSLAVAKIIIEFNPMYPKNDIALMKLOPPLTFS 298
QY 302 GTVRPICLPFFDEELTPATPLWIIIGWFTKQNGKMSDILLQASVQVIDSTRCNADDA 361
DB 299 GTVRPICLPFFDEELTPATPLWIIIGWFTKQNGKMSDILLQASVQVIDSTRCNADDA 358
QY 362 GEVTEKMCAGIPGEGVDTCQDSGGPLMYQSDQHVGVISWGYCGGPGSTPGVYTKVS 421
DB 359 GEVTEKMCAGIPGEGVDTCQDSGGPLMYQSDQHVGVISWGYCGGPGSTPGVYTKVS 418
QY 422 AYLNWIYNNWKAEL 435
DB 419 AYLNWIYNNWKAEL 432

RESULT 7
US-10-063-735-112

RESULT 6
US-10-015-387A-275
Sequence 275, Application US/10015387A
Publication No. US20030135034A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Botstein, David
APPLICANT: Desnoyers, Luc
APPLICANT: Eaton, Dan I.
APPLICANT: Ferrara, Napoleone
APPLICANT: Fong, Sherman
APPLICANT: Gao, Wei-Qiang
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Grimaldi, Christopher J.
APPLICANT: Gurney, Austin L.
APPLICANT: Hillan, Kenneth J.
APPLICANT: Pan, James
APPLICANT: Paoni, Nicholas F.
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
FILE OF INVENTION: Acids Encoding the Same
FILE REFERENCE: P2830PIC54
CURRENT APPLICATION NUMBER: US/10/015,387A
CURRENT FILING DATE: 2001-12-12
Prior Application removed - See File Wrapper or Palm
NUMBER OF SEQ ID NOS: 477
SEQ ID NO 275
LENGTH: 432
TYPE: PRT
ORGANISM: Homo sapiens
US-10-015-387A-275

Query Match 98.1%; Score 2297.5; DB 12; Length 432;
Best Local Similarity 98.8%; Pred. No. 6.1e-222;
Matches 429; Conservative 0; Mismatches 0; Indels 5; Gaps 1;
QY 2 DPDSQPLNSLDVKPRIPMETFRKVGIPPIIIALLSLASIIIVVLIKVLIDKYFL 61
DB 4 DPDSQPLNSLDVKPRIPMETFRKVGIPPIIIALLSLASIIIVVLIKVLIDKYFL 63
QY 62 CGQPLHFIPRKQDCGELDCPLGEDEBEHCVKSPPEGPAVAVRLSKDRSTLQVLDSATGNW 121
DB 64 CGQPLHFIPRKQDCGELDCPLGEDEBEHCVKSPPEGPAVAVRLSKDRSTLQVLDSATGNW 123
QY 122 FSACFDNFTALAEACRQMGYSKPTTFAVEIGPDQDLDVVEITENSQELMRNSSGFC 181
DB 124 FSACFDNFTALAEACRQMGYSKPTTFAVEIGPDQDLDVVEITENSQELMRNSSGFC 178
QY 182 LSGSLVSLHCLACGKSLKTPRVVGGEEASVDSWPQVSIQYDKQHVCGGSIIDPHWVLT 241
DB 179 LSGSLVSLHCLACGKSLKTPRVVGGEEASVDSWPQVSIQYDKQHVCGGSIIDPHWVLT 238
QY 242 AHCPRKHTDVFNNKVRAGSKLGSPFSLAVAKIIIEFNPMYPKNDIALMKLOPPLTFS 301
DB 239 AHCPRKHTDVFNNKVRAGSKLGSPFSLAVAKIIIEFNPMYPKNDIALMKLOPPLTFS 298
QY 302 GTVRPICLPFFDEELTPATPLWIIIGWFTKQNGKMSDILLQASVQVIDSTRCNADDA 361
DB 299 GTVRPICLPFFDEELTPATPLWIIIGWFTKQNGKMSDILLQASVQVIDSTRCNADDA 358
QY 362 GEVTEKMCAGIPGEGVDTCQDSGGPLMYQSDQHVGVISWGYCGGPGSTPGVYTKVS 421
DB 359 GEVTEKMCAGIPGEGVDTCQDSGGPLMYQSDQHVGVISWGYCGGPGSTPGVYTKVS 418
QY 422 AYLNWIYNNWKAEL 435
DB 419 AYLNWIYNNWKAEL 432

Sequence 112, Application US/10063735
Publication No. US20030138882A1

GENERAL INFORMATION:

APPLICANT: Eaton, Dan L.

APPLICANT: Filvaroff, Ellen

APPLICANT: Gerritsen, Mary E.

APPLICANT: Goddard, Audrey

APPLICANT: Godowski, Paul J.

APPLICANT: Grimaldi, Christopher J.

APPLICANT: Gurney, Austin L.

APPLICANT: Watanabe, Colin K.

APPLICANT: Wood, William I.

TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC

FILE REFERENCE: F3230R1C1

CURRENT APPLICATION NUMBER: US/10/063,735

Prior Application removed - See Palm or File Wrapper

NUMBER OF SEQ ID NOS: 170

SEQ ID NO 112

LENGTH: 432

TYPE: PRT

ORGANISM: Homo Sapien

US-10-063-735-112

Query Match 98.1%; Score 2297.5; DB 12; Length 432;
Best Local Similarity 98.8%; Pred. No. 6.1e-222;

Matches 429; Conservative 0; Mismatches 0; Indels 5; Gaps 1;

QY 2 DPDSQPLNSLDVVKPLRKPRIPMETFRKVGIPPIIIALLSLASIIIVVLLIKVILDKYIFL 61

Db 4 DPDSQPLNSLDVVKPLRKPRIPMETFRKVGIPPIIIALLSLASIIIVVLLIKVILDKYIFL 63

QY 62 CGQPLHFIPRKQLCDGELDCPLGDEDEHCVKSPPEGPAVAVRLSKDRSTLQVLDSATGNW 121

Db 64 CGQPLHFIPRKQLCDGELDCPLGDEDEHCVKSPPEGPAVAVRLSKDRSTLQVLDSATGNW 123

QY 122 FSACFNFTALAEATACRQMGYSKPTFRAVEIGPDODLDVVEITENSQBLRNSSGPGC 181

Db 124 FSACFNFTALAEATACRQMGYS-----RAVEIGPDODLDVVEITENSQBLRNSSGPGC 178

QY 182 LSGSLVSLHCLACGSKLTPRVVVGEEASVDSWPMQVSIQYDKQHVCGGSIIDPHWVLT 241

Db 179 LSGSLVSLHCLACGSKLTPRVVVGEEASVDSWPMQVSIQYDKQHVCGGSIIDPHWVLT 238

QY 242 AHCFRKHDTFVNKVRAGSKLGSFPPLAVAKIIIEFNPMYPKNDIALMKLQFPLTFS 301

Db 239 AHCFRKHDTFVNKVRAGSKLGSFPPLAVAKIIIEFNPMYPKNDIALMKLQFPLTFS 298

QY 302 GTVRPCLPFPDEELTPATPLWIIIGWFTKQNGKMSDILLOASVQVIDSTRCNADDAQ 361

Db 299 GTVRPCLPFPDEELTPATPLWIIIGWFTKQNGKMSDILLOASVQVIDSTRCNADDAQ 358

QY 362 GEVTERKMCAGIPEGGVDTCCQDSGGPLMYQSDQHWVGVIVSWGYCGGSPSTPGVYTKVS 421

Db 359 GEVTERKMCAGIPEGGVDTCCQDSGGPLMYQSDQHWVGVIVSWGYCGGSPSTPGVYTKVS 418

QY 422 AYLNWIYVNWKAEL 435

Db 419 AYLNWIYVNWKAEL 432

RESULT 8

US-10-006-130A-275

Sequence 275, Application US/10006130A

Publication No. US20030148375A1

GENERAL INFORMATION:

APPLICANT: Baker, Kevin P.

APPLICANT: Botstein, David

APPLICANT: Desnoyers, Luc

APPLICANT: Eaton, Dan I.

APPLICANT: Ferrara, Napoleone

APPLICANT: Fong, Sherman

APPLICANT: Gao, Wei-Qiang
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Grimaldi, Christopher J.
APPLICANT: Gurney, Austin L.
APPLICANT: Hillan, Kenneth J.
APPLICANT: Pan, James
APPLICANT: Paoni, Nicholas F.
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
FILE REFERENCE: P2830PIC7
CURRENT APPLICATION NUMBER: US/10/006,130A
CURRENT FILING DATE: 2002-03-19
Prior Application removed - See File Wrapper or Palm
NUMBER OF SEQ ID NOS: 477
SEQ ID NO 275
LENGTH: 432
TYPE: PRT
ORGANISM: Homo sapiens
US-10-006-130A-275

Query Match 98.1%; Score 2297.5; DB 12; Length 432;
Best Local Similarity 98.8%; Pred. No. 6.1e-222;

Matches 429; Conservative 0; Mismatches 0; Indels 5; Gaps 1;

QY 2 DPDSQPLNSLDVVKPLRKPRIPMETFRKVGIPPIIIALLSLASIIIVVLLIKVILDKYIFL 61

Db 4 DPDSQPLNSLDVVKPLRKPRIPMETFRKVGIPPIIIALLSLASIIIVVLLIKVILDKYIFL 63

QY 62 CGQPLHFIPRKQLCDGELDCPLGDEDEHCVKSPPEGPAVAVRLSKDRSTLQVLDSATGNW 121

Db 64 CGQPLHFIPRKQLCDGELDCPLGDEDEHCVKSPPEGPAVAVRLSKDRSTLQVLDSATGNW 123

QY 122 FSACFNFTALAEATACRQMGYSKPTFRAVEIGPDODLDVVEITENSQBLRNSSGPGC 181

Db 124 FSACFNFTALAEATACRQMGYS-----RAVEIGPDODLDVVEITENSQBLRNSSGPGC 178

QY 182 LSGSLVSLHCLACGSKLTPRVVVGEEASVDSWPMQVSIQYDKQHVCGGSIIDPHWVLT 241

Db 179 LSGSLVSLHCLACGSKLTPRVVVGEEASVDSWPMQVSIQYDKQHVCGGSIIDPHWVLT 238

QY 242 AHCFRKHDTFVNKVRAGSKLGSFPPLAVAKIIIEFNPMYPKNDIALMKLQFPLTFS 301

Db 239 AHCFRKHDTFVNKVRAGSKLGSFPPLAVAKIIIEFNPMYPKNDIALMKLQFPLTFS 298

QY 302 GTVRPCLPFPDEELTPATPLWIIIGWFTKQNGKMSDILLOASVQVIDSTRCNADDAQ 361

Db 299 GTVRPCLPFPDEELTPATPLWIIIGWFTKQNGKMSDILLOASVQVIDSTRCNADDAQ 358

QY 362 GEVTERKMCAGIPEGGVDTCCQDSGGPLMYQSDQHWVGVIVSWGYCGGSPSTPGVYTKVS 421

Db 359 GEVTERKMCAGIPEGGVDTCCQDSGGPLMYQSDQHWVGVIVSWGYCGGSPSTPGVYTKVS 418

QY 422 AYLNWIYVNWKAEL 435

Db 419 AYLNWIYVNWKAEL 432

RESULT 9

US-10-199-672-330

Sequence 330, Application US/10199672

Publication No. US20030148442A1

GENERAL INFORMATION:

APPLICANT: Baker, Kevin P.

APPLICANT: Chen, Jian

APPLICANT: Desnoyers, Luc

APPLICANT: Goddard, Audrey

APPLICANT: Godowski, Paul J.

APPLICANT: Gurney, Austin L.

APPLICANT: Pan, James

APPLICANT: Smith, Victoria

APPLICANT: Watanabe, Colin K.

APPLICANT: Wood, William I.

```

; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C1
; CURRENT APPLICATION NUMBER: US/10/199,672
; CURRENT FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: US/10/052,586
; PRIOR FILING DATE: 2002-01-15
; PRIOR APPLICATION NUMBER: 60/059263
; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 60/059266
; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 60/062250
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/063120
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/063121
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/063486
; PRIOR FILING DATE: 1997-10-21
; PRIOR APPLICATION NUMBER: 60/063540
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/063541
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/063544
; PRIOR FILING DATE: 1997-10-28
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 330
; LENGTH: 432
; TYPE: PRT
; ORGANISM: Homo Sapien
; US-10-199-672-330

Query Match      98.1%; Score 2297.5; DB 12; Length 432;
Best Local Similarity 98.8%; Pred. No. 6.1e-222;
Matches 429; Conservative 0; Mismatches 0; Indels 5; Gaps 1;

QY      2  DPDSQPLNSLDVKPIKRPINPMPETFRKVGIPPIIIALLSLASIIIVVLIKVLIDKYFL 61
Db      4  DPDSQPLNSLDVKPIKRPINPMPETFRKVGIPPIIIALLSLASIIIVVLIKVLIDKYFL 63
QY      62  CGQPLHFI PRKQLCDGELDCPLGCEDEEHCVKSPPEGPAVAVRLSKDRSTLQVLDSATGNW 121
Db      64  CGQPLHFI PRKQLCDGELDCPLGCEDEEHCVKSPPEGPAVAVRLSKDRSTLQVLDSATGNW 123
QY      122  FSACFNFTEALAEACRQMGYSKPTFFRAVEIGPDQDLVVEITENSQELRNSSGPC 181
Db      124  FSACFNFTEALAEACRQMGYS-----RAVEIGPDQDLVVEITENSQELRNSSGPC 178
QY      182  LSGSLVSLHCLAGKSLKTRPVVGGSEASVDSWPQVSIQYDKQHVCGGSILDPHWLTA 241
Db      179  LSGSLVSLHCLAGKSLKTRPVVGGSEASVDSWPQVSIQYDKQHVCGGSILDPHWLTA 238
QY      242  AHCFRKHTDVFNKVRAGSKLSPFSLAVAKIIIEFNPMYKNDIALMKLOFPLTFS 301
Db      239  AHCFRKHTDVFNKVRAGSKLSPFSLAVAKIIIEFNPMYKNDIALMKLOFPLTFS 298
QY      302  GTVRPICLFFDEELPATPLWTIIGFTKQNGKMSDILLQASVQVIDSTRCNADDAYQ 361
Db      299  GTVRPICLFFDEELPATPLWTIIGFTKQNGKMSDILLQASVQVIDSTRCNADDAYQ 358
QY      362  GEVTEKQMCAGIEGGVDTCQSGGGLPMYQSDQWVGVISGCGSPSTEGVTKYS 421
Db      359  GEVTEKQMCAGIEGGVDTCQSGGGLPMYQSDQWVGVISGCGSPSTEGVTKYS 418
QY      422  AYLNIWNIYNWKAEI 435
Db      419  AYLNIWNIYNWKAEI 432

RESULT 10
US-10-006-172A-275
; Sequence 275, Application US/10006172A
; Publication No. US20030153000A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan I.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth J.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE REFERENCE: P2830PIC11
; CURRENT APPLICATION NUMBER: US/10/006,172A
; CURRENT FILING DATE: 2002-03-19
; PRIOR APPLICATION NUMBER: 60/098716
; PRIOR FILING DATE: 1998-09-01
; PRIOR APPLICATION NUMBER: 60/098723
; PRIOR FILING DATE: 1998-09-01
; PRIOR APPLICATION NUMBER: 60/098749
; PRIOR FILING DATE: 1998-09-01
; PRIOR APPLICATION NUMBER: 60/098750
; PRIOR FILING DATE: 1998-09-01
; PRIOR APPLICATION NUMBER: 60/098803
; PRIOR FILING DATE: 1998-09-02
; PRIOR APPLICATION NUMBER: 60/098821
; PRIOR FILING DATE: 1998-09-02
; PRIOR APPLICATION NUMBER: 60/098843
; PRIOR FILING DATE: 1998-09-02
; PRIOR APPLICATION NUMBER: 60/099536
; PRIOR FILING DATE: 1998-09-09
; PRIOR APPLICATION NUMBER: 60/099596
; PRIOR FILING DATE: 1998-09-09
; PRIOR APPLICATION NUMBER: 60/099598
; PRIOR FILING DATE: 1998-09-09
; PRIOR APPLICATION NUMBER: 60/099602
; PRIOR FILING DATE: 1998-09-09
; PRIOR APPLICATION NUMBER: 60/099642
; PRIOR FILING DATE: 1998-09-09
; PRIOR APPLICATION NUMBER: 60/099741
; PRIOR FILING DATE: 1998-09-10
; PRIOR APPLICATION NUMBER: 60/099754
; PRIOR FILING DATE: 1998-09-10
; PRIOR APPLICATION NUMBER: 60/099763
; PRIOR FILING DATE: 1998-09-10
; PRIOR APPLICATION NUMBER: 60/099792
; PRIOR FILING DATE: 1998-09-10
; PRIOR APPLICATION NUMBER: 60/099808
; PRIOR FILING DATE: 1998-09-10
; PRIOR APPLICATION NUMBER: 60/099812
; PRIOR FILING DATE: 1998-09-10
; PRIOR APPLICATION NUMBER: 60/099815
; PRIOR FILING DATE: 1998-09-10
; PRIOR APPLICATION NUMBER: 60/099816
; PRIOR FILING DATE: 1998-09-10
; PRIOR APPLICATION NUMBER: 60/100385
; PRIOR FILING DATE: 1998-09-15
; PRIOR APPLICATION NUMBER: 60/100388
; PRIOR FILING DATE: 1998-09-15
; PRIOR APPLICATION NUMBER: 60/100390
; PRIOR FILING DATE: 1998-09-15
; PRIOR APPLICATION NUMBER: 60/100584
; PRIOR FILING DATE: 1998-09-16
; PRIOR APPLICATION NUMBER: 60/100627
; PRIOR FILING DATE: 1998-09-16
; PRIOR APPLICATION NUMBER: 60/100661
```


1	PRIOR APPLICATION NUMBER: 60/102687
2	PRIOR FILING DATE: 1998-10-01
3	PRIOR APPLICATION NUMBER: 60/102965
4	PRIOR FILING DATE: 1998-10-02
5	PRIOR APPLICATION NUMBER: 60/103258
6	PRIOR FILING DATE: 1998-10-06
7	PRIOR APPLICATION NUMBER: 60/103314
8	PRIOR FILING DATE: 1998-10-07
9	PRIOR APPLICATION NUMBER: 60/103315
10	PRIOR FILING DATE: 1998-10-07
11	PRIOR APPLICATION NUMBER: 60/103328
12	PRIOR FILING DATE: 1998-10-07
13	PRIOR APPLICATION NUMBER: 60/103395
14	PRIOR FILING DATE: 1998-10-07
15	PRIOR APPLICATION NUMBER: 60/103396
16	PRIOR FILING DATE: 1998-10-07
17	PRIOR APPLICATION NUMBER: 60/103401
18	PRIOR FILING DATE: 1998-10-07
19	PRIOR APPLICATION NUMBER: 60/103449
20	PRIOR FILING DATE: 1998-10-06
21	PRIOR APPLICATION NUMBER: 60/103633
22	PRIOR FILING DATE: 1998-10-08
23	PRIOR APPLICATION NUMBER: 60/103678
24	PRIOR FILING DATE: 1998-10-08
25	PRIOR APPLICATION NUMBER: 60/103679
26	PRIOR FILING DATE: 1998-10-08
27	PRIOR APPLICATION NUMBER: 60/103711
28	PRIOR FILING DATE: 1998-10-08
29	PRIOR APPLICATION NUMBER: 60/104257
30	PRIOR FILING DATE: 1998-10-14
31	PRIOR APPLICATION NUMBER: 60/104987
32	PRIOR FILING DATE: 1998-10-20
33	PRIOR APPLICATION NUMBER: 60/105000
34	PRIOR FILING DATE: 1998-10-20
35	PRIOR APPLICATION NUMBER: 60/105002
36	PRIOR FILING DATE: 1998-10-20
37	PRIOR APPLICATION NUMBER: 60/105104
38	PRIOR FILING DATE: 1998-10-21
39	PRIOR APPLICATION NUMBER: 60/105169
40	PRIOR FILING DATE: 1998-10-22
41	PRIOR APPLICATION NUMBER: 60/105266
42	PRIOR FILING DATE: 1998-10-22
43	PRIOR APPLICATION NUMBER: 60/105693
44	PRIOR FILING DATE: 1998-10-26
45	PRIOR APPLICATION NUMBER: 60/105694
46	PRIOR FILING DATE: 1998-10-26
47	PRIOR APPLICATION NUMBER: 60/105807
48	PRIOR FILING DATE: 1998-10-27
49	PRIOR APPLICATION NUMBER: 60/105881
50	PRIOR FILING DATE: 1998-10-27
51	PRIOR APPLICATION NUMBER: 60/105882
52	PRIOR FILING DATE: 1998-10-27
53	PRIOR APPLICATION NUMBER: 60/106023
54	PRIOR FILING DATE: 1998-10-28

Best Local Similarity 98.8%; Pred. No. 6.1e-222;

QY	2	DPDSQPLNSLDVKPLRPKPRIPMETFRKVGPIIIALLSLASIIIVVVLKVIDLKKYFL	61
DB	4	DPDSQPLNSLDVKPLRPKPRIPMETFRKVGPIIIALLSLASIIIVVVLKVIDLKKYFL	63
QY	62	CQCPHLHFIKRLCDGELDCPLGEDDEEHCVKSPFEGPAVAVRLSKDRSTLQVLDSATGNW	12
DB	64	CQCPHLHFIKRLCDGELDCPLGEDDEEHCVKSPFEGPAVAVRLSKDRSTLQVLDSATGNW	12
QY	122	FSACDNTFTEALAEATACRQMGYSKPTTPEAVEIGDPQDLDVETENSQELMRNNSGPC	18
DB	124	FSACDNTFTEALAEATACRQMGYS-----RAVEIGDPQDLDVETENSQELMRNNSGPC	17
QY	182	LSGSLVSLHLCLACGSKSLKTPRVVGGEASVDSWPQVSIQYDKHVCVCGSLIDPHWLTA	24

Db 179 LSGSLVSLHCLACGKSLKTRVVGGEASVDSWPMQVSIQYDKQHVCGGSLDPHVLTA 238
QY 242 AHCERKHTDVFNKVRAGSKLGFPSLAVAKIIIEFNPMYPKNDIALMKLQFPLTFS 301
Db 239 AHCERKHTDVFNKVRAGSKLGFPSLAVAKIIIEFNPMYPKNDIALMKLQFPLTFS 298
QY 302 GTVRPCLPFDEBELTPATLWIIIGWFTKONGGKMSDILLOASVQVIDSTRCNADAYQ 361
Db 299 GTVRPCLPFDEBELTPATLWIIIGWFTKONGGKMSDILLOASVQVIDSTRCNADAYQ 358
QY 362 GEVTERKMCAGIPEGGVDTCQGSGGPLMYQSDQVHVGIVSWGVCSTEGVYTKVS 421
Db 359 GEVTERKMCAGIPEGGVDTCQGSGGPLMYQSDQVHVGIVSWGVCSTEGVYTKVS 418
QY 422 AYLNWIYNNWKAEL 435
Db 419 AYLNWIYNNWKAEL 432

RESULT 11

US-10-187-749-330
; Sequence 330, Application US/10187749
; Publication No. US20030153036A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C1
; CURRENT APPLICATION NUMBER: US/10/187,749
; CURRENT FILING DATE: 2002-07-01
; PRIOR APPLICATION NUMBER: US/10/052,586
; PRIOR FILING DATE: 2002-01-15
; PRIOR APPLICATION NUMBER: 60/059263
; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 60/059266
; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 60/062250
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/063120
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/063121
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/063486
; PRIOR FILING DATE: 1997-10-21
; PRIOR APPLICATION NUMBER: 60/063540
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/063541
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/063544
; PRIOR FILING DATE: 1997-10-28
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 330
; LENGTH: 432
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-187-749-330
Query Match 98.1%; Score 2297.5; DB 12; Length 432;
Best Local Similarity 98.8%; Pred. No. 6.1e-222;
Matches 429; Conservative 0; Mismatches 0; Indels 5; Gaps 1;
QY 2 DPDSQPLNSLDVKKPRKPIPMETFRKVGIPITIIALLSLASIIIVVLKILDKYIFL 61

Db 4 DPDSQPLNSLDVKKPRKPIPMETFRKVGIPITIIALLSLASIIIVVLKILDKYIFL 63
QY 62 CQOPLHFIKRCQDGLDCPLGEDEEHCVKGFPEGPAVAVRLSKDRSTLQVLDSATGNW 121
Db 64 CQOPLHFIKRCQDGLDCPLGEDEEHCVKGFPEGPAVAVRLSKDRSTLQVLDSATGNW 123
QY 122 FSACFDNFTALAEACRQMGYSKPTFRVIGIPDODLDVVEITENSQELRNSSGPG 181
Db 124 FSACFDNFTALAEACRQMGYS-----RAVIGIPDODLDVVEITENSQELRNSSGPG 178
QY 182 LSGSLVSLHCLACGKSLKTRVVGGEASVDSWPMQVSIQYDKQHVCGGSLDPHVLTA 241
Db 179 LSGSLVSLHCLACGKSLKTRVVGGEASVDSWPMQVSIQYDKQHVCGGSLDPHVLTA 238
QY 242 AHCERKHTDVFNKVRAGSKLGFPSLAVAKIIIEFNPMYPKNDIALMKLQFPLTFS 301
Db 239 AHCERKHTDVFNKVRAGSKLGFPSLAVAKIIIEFNPMYPKNDIALMKLQFPLTFS 298
QY 302 GTVRPCLPFDEBELTPATLWIIIGWFTKONGGKMSDILLOASVQVIDSTRCNADAYQ 361
Db 299 GTVRPCLPFDEBELTPATLWIIIGWFTKONGGKMSDILLOASVQVIDSTRCNADAYQ 358
QY 362 GEVTERKMCAGIPEGGVDTCQGSGGPLMYQSDQVHVGIVSWGVCSTEGVYTKVS 421
Db 359 GEVTERKMCAGIPEGGVDTCQGSGGPLMYQSDQVHVGIVSWGVCSTEGVYTKVS 418
QY 422 AYLNWIYNNWKAEL 435
Db 419 AYLNWIYNNWKAEL 432

RESULT 12

US-10-194-457-330
; Sequence 330, Application US/10194457
; Publication No. US20030153037A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C296
; CURRENT APPLICATION NUMBER: US/10/194,457
; CURRENT FILING DATE: 2002-07-11
; PRIOR APPLICATION NUMBER: 10/052586
; PRIOR FILING DATE: 2002-01-15
; PRIOR APPLICATION NUMBER: 60/059263
; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 60/059266
; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 60/062250
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/063120
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/063121
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/063486
; PRIOR FILING DATE: 1997-10-21
; PRIOR APPLICATION NUMBER: 60/063540
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/063541
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/063544
; PRIOR FILING DATE: 1997-10-28

;; Prior Application data removed - See File Wrapper or PALM.

;; NUMBER OF SEQ ID NOS: 612

;; Best Local Similarity 98.8%; Pred. No. 6.1e-222;

Matches 429; Conservative 0; Mismatches 0; Indels 5; Gaps 1;

;; LENGTH: 432

;; TYPE: PRT

;; ORGANISM: Homo Sapien

US-10-194-457-330

Query Match 98.1%; Score 2297.5; DB 12; Length 432;
Best Local Similarity 98.8%; Pred. No. 6.1e-222;
Matches 429; Conservative 0; Mismatches 0; Indels 5; Gaps 1;
QY 2 DPDSQPLNSLDVPLKRPRIIPMETFRKVGPIIPIIALLSLASIIIVVVLKVLIDKYFL 61
Db 4 DPDSQPLNSLDVPLKRPRIIPMETFRKVGPIIPIIALLSLASIIIVVVLKVLIDKYFL 63
QY 62 CGOPLHFIIPKQLCDGELDCPLGEDEEHCVKSPFEGPAVAVRLSKDRSTIQVLD SATGNW 121
Db 64 CGOPLHFIIPKQLCDGELDCPLGEDEEHCVKSPFEGPAVAVRLSKDRSTIQVLD SATGNW 123
QY 122 FSACFNFTALAEATACRQMGYSKPTFRAVEIGPDQDLDVVEITENSQELMRNSSGPC 181
Db 124 FSACFNFTALAEATACRQMGYS-----RAVEIGPDQDLDVVEITENSQELMRNSSGPC 178
QY 182 LSGSLVSLHCLACGSKLTPRVVGGEEASVDSWPQVSIQYDKQHVCGGSIIDPHWVLT 241
Db 179 LSGSLVSLHCLACGSKLTPRVVGGEEASVDSWPQVSIQYDKQHVCGGSIIDPHWVLT 238
QY 242 AHCFRKHDTVFNNKVRAGSKLSPFSLAVAKIIIEFNPMYPKNDIALMKLOFPLTFS 301
Db 239 AHCFRKHDTVFNNKVRAGSKLSPFSLAVAKIIIEFNPMYPKNDIALMKLOFPLTFS 298
QY 302 GTVRPICLPFDEELTPATPLMIIGWFTKQNGKMSDILLQASVQVIDSTRCNADAY 361
Db 299 GTVRPICLPFDEELTPATPLMIIGWFTKQNGKMSDILLQASVQVIDSTRCNADAY 358
QY 362 GEVTEKMCAGIPEGVDTCQDSSGGPLMYQSDQHVGVISWGYCGGSPSTPGVTKVS 421
Db 359 GEVTEKMCAGIPEGVDTCQDSSGGPLMYQSDQHVGVISWGYCGGSPSTPGVTKVS 418
QY 422 AYLNWIYNNWKAE 435
Db 419 AYLNWIYNNWKAE 432

RESULT 13

US-10-184-642-330

;; Sequence 330, Application US/10184642

;; Publication No. US20030157635A1

;; GENERAL INFORMATION:

;; APPLICANT: Baker, Kevin P.

;; APPLICANT: Chen, Jian

;; APPLICANT: Desnoyers, Luc

;; APPLICANT: Goddard, Audrey

;; APPLICANT: Godowski, Paul J.

;; APPLICANT: Gurney, Austin L.

;; APPLICANT: Pan, James

;; APPLICANT: Smith, Victoria

;; APPLICANT: Watanabe, Colin K.

;; APPLICANT: Wood, William I.

;; APPLICANT: Zhang, Zemin

;; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC

;; FILE REFERENCE: P3430R1C194

;; CURRENT APPLICATION NUMBER: US/10/184,642

;; CURRENT FILING DATE: 2002-06-27

;; Prior Application removed - See File Wrapper or Palm

;; NUMBER OF SEQ ID NOS: 612

;; SEQ ID NO 330

;; LENGTH: 432

;; TYPE: PRT

;; ORGANISM: Homo Sapien

US-10-184-642-330

Query Match 98.1%; Score 2297.5; DB 12; Length 432;

Best Local Similarity 98.8%; Pred. No. 6.1e-222;

Matches 429; Conservative 0; Mismatches 0; Indels 5; Gaps 1;

QY 2 DPDSQPLNSLDVPLKRPRIIPMETFRKVGPIIPIIALLSLASIIIVVVLKVLIDKYFL 61

Db 4 DPDSQPLNSLDVPLKRPRIIPMETFRKVGPIIPIIALLSLASIIIVVVLKVLIDKYFL 63

QY 62 CGOPLHFIIPKQLCDGELDCPLGEDEEHCVKSPFEGPAVAVRLSKDRSTIQVLD SATGNW 121

Db 64 CGOPLHFIIPKQLCDGELDCPLGEDEEHCVKSPFEGPAVAVRLSKDRSTIQVLD SATGNW 123

QY 122 FSACFNFTALAEATACRQMGYSKPTFRAVEIGPDQDLDVVEITENSQELMRNSSGPC 181

Db 124 FSACFNFTALAEATACRQMGYS-----RAVEIGPDQDLDVVEITENSQELMRNSSGPC 178

QY 182 LSGSLVSLHCLACGSKLTPRVVGGEEASVDSWPQVSIQYDKQHVCGGSIIDPHWVLT 241

Db 179 LSGSLVSLHCLACGSKLTPRVVGGEEASVDSWPQVSIQYDKQHVCGGSIIDPHWVLT 238

QY 242 AHCFRKHDTVFNNKVRAGSKLSPFSLAVAKIIIEFNPMYPKNDIALMKLOFPLTFS 301

Db 239 AHCFRKHDTVFNNKVRAGSKLSPFSLAVAKIIIEFNPMYPKNDIALMKLOFPLTFS 298

QY 302 GTVRPICLPFDEELTPATPLMIIGWFTKQNGKMSDILLQASVQVIDSTRCNADAY 361

Db 299 GTVRPICLPFDEELTPATPLMIIGWFTKQNGKMSDILLQASVQVIDSTRCNADAY 358

QY 362 GEVTEKMCAGIPEGVDTCQDSSGGPLMYQSDQHVGVISWGYCGGSPSTPGVTKVS 421

Db 359 GEVTEKMCAGIPEGVDTCQDSSGGPLMYQSDQHVGVISWGYCGGSPSTPGVTKVS 418

QY 422 AYLNWIYNNWKAE 435

Db 419 AYLNWIYNNWKAE 432

RESULT 14

US-10-196-747-330

;; Sequence 330, Application US/10196747

;; Publication No. US20030162250A1

;; GENERAL INFORMATION:

;; APPLICANT: Baker, Kevin P.

;; APPLICANT: Chen, Jian

;; APPLICANT: Desnoyers, Luc

;; APPLICANT: Goddard, Audrey

;; APPLICANT: Godowski, Paul J.

;; APPLICANT: Gurney, Austin L.

;; APPLICANT: Pan, James

;; APPLICANT: Smith, Victoria

;; APPLICANT: Watanabe, Colin K.

;; APPLICANT: Wood, William I.

;; APPLICANT: Zhang, Zemin

;; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC

;; FILE REFERENCE: P3430R1C346

;; CURRENT APPLICATION NUMBER: US/10/196,747

;; CURRENT FILING DATE: 2002-07-16

;; Prior Application removed - See File Wrapper or Palm

;; NUMBER OF SEQ ID NOS: 612

;; SEQ ID NO 330

;; LENGTH: 432

;; TYPE: PRT

;; ORGANISM: Homo Sapien

US-10-196-747-330

Query Match 98.1%; Score 2297.5; DB 12; Length 432;

Best Local Similarity 98.8%; Pred. No. 6.1e-222;

Matches 429; Conservative 0; Mismatches 0; Indels 5; Gaps 1;

QY 2 DPDSQPLNSLDVPLKRPRIIPMETFRKVGPIIPIIALLSLASIIIVVVLKVLIDKYFL 61

Result No.	Score	Query Match	Length	DB	ID	Description
1	584.5	25.0	417	1	S00845	hepsin (EC 3.4.21.
2	582.5	24.9	1019	1	A56318	enteropeptidase (E
3	571	24.4	1035	1	A43090	enteropeptidase (E
4	568.5	24.3	416	1	S33777	hepsin (EC 3.4.21.
5	552	23.6	1034	1	A53663	enteropeptidase (E
6	511	21.8	638	1	KQKRLP	plasma kallikrein
7	508	21.7	638	1	KQMSPL	plasma kallikrein
8	505	21.6	436	2	JX0172	acrosin (EC 3.4.21
9	504.5	21.5	1524	2	T30337	polyprotein - Afri
10	503	21.5	418	2	A37344	acrosin (EC 3.4.21
11	501.5	21.4	437	2	S18407	acrosin (EC 3.4.21
12	493.5	21.1	421	2	S23599	acrosin (EC 3.4.21
13	491.5	21.0	638	1	KQKRP	plasma kallikrein
14	488.5	20.9	1113	2	JE0315	low-density lipopr
15	481	20.5	421	1	S11674	acrosin (EC 3.4.21
16	478.5	20.4	415	1	A34170	acrosin (EC 3.4.21
17	474	20.2	855	2	JC7731	membrane-bound arg
18	470	20.1	761	2	JCS759	brain-specific ser
19	468.5	20.0	431	2	S47538	acrosin (EC 3.4.21
20	461.5	19.7	277	2	S35340	trypsin (EC 3.4.21
21	460.5	19.7	263	1	KYRTB	chymotrypsin (EC 3
22	458.5	19.6	267	2	S40006	trypsin (EC 3.4.21
23	458.5	19.6	790	1	PLPG	plasma (EC 3.4.21
24	457.5	19.5	625	1	KEHUL	coagulation factor
25	455.5	19.4	812	1	PLMS	plasmin (EC 3.4.21
26	449	19.2	420	2	A55283	acrosin (EC 3.4.21
27	448	19.1	810	2	I46260	plasmin (EC 3.4.21
28	447.5	19.1	343	1	A57014	prostatin (EC 3.4.
29	445.5	19.0	455	2	A61545	plasmin (EC 3.4.21

```

QY 354 CNADAYQGEVTEKMMKAGIPGGVDTCQDSGGPLMYQ-----SQDWHVGVISWGYGC 408
Db 322 CNGADFYGQIKPKMFCAGYPEGGIDACQDSGGPFVCDSDISRTFRWLGGIVSWGTC 381
QY 409 GGPSTPGVYTKVSAYLNWYNWK 432
Db 382 ALAQKPGVYTKVSDFREWFQAIK 405

RESULT 2
A56318
enteropeptidase (EC 3.4.21.9) precursor [validated] - human
N:Alternate names: enterokinase
C:Species: Homo sapiens (man)
C:Date: 19-May-1995 #sequence_revision 09-Aug-1996 #text_change 28-Apr-2003
C:Accession: A56318; B43090
R:Kitamoto, Y.; Velle, R.A.; Donis-Keller, H.; Sadler, J.E.
Biochemistry 34, 4562-4568, 1995
A:Title: cDNA sequence and chromosomal localization of human enterokinase, the proteolytic
A:Reference number: A56318; MUID:95234679; PMID:7718557
A:Accession: A56318
A:Molecule type: mRNA
A:Residues: 1-1019 <KIT>
A:Cross-references: GB:U09860; NID:g746412; PIDN:AAC50138.1; PID:g746413
R:Kitamoto, Y.; Yuan, X.; Wu, Q.; McCourt, D.W.; Sadler, J.E.
Proc. Natl. Acad. Sci. U.S.A. 91, 7588-7592, 1994
A:Title: Enterokinase, the initiator of intestinal digestion, is a mosaic protease composed
A:Reference number: A43090; MUID:94329561; PMID:8052624
A:Accession: B43090
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 749-1019 <KIT>
A:Cross-references: GB:U09860
C:Comment: The mechanism of association with the membrane of the intestinal brush border
created below) or with amino-terminal myristoylation of the heavy chain.
C:Genetics:
A:Gene: GDB:FRSS7
A:Cross-references: GDB:384083; OMIM:226200
A:Map position: 21q21-21q21
C:Complex: Mature enteropeptidase is variously reported to contain two (heavy and light)
ed by a disulfide bond. Possibly, conversion from membrane-bound to soluble forms involve
ducts.
C:Function:
A:Description: cleaves activation peptide from trypsinogen to produce active trypsin
A:Pathway: intestinal digestive hydrolase cascade
C:Superfamily: enteropeptidase; C1r/C1s repeat homology; LDL receptor ligand-binding rep
C:Keywords: glycoprotein; hydrolase; serine proteinase; transmembrane protein; zymogen
F:1-784/Product: enteropeptidase heavy chain #status predicted <TM>
F:22-38/Domain: transmembrane #status predicted <TM>
F:184-221/Domain: LDL receptor ligand-binding repeat homology <LDL1>
F:342-504/Domain: NAM homology <NAM>
F:526-631/Domain: C1r/C1s repeat homology <C1R>
F:643-677/Domain: LDL receptor ligand-binding repeat homology <LDL2>
F:678-783/Domain: scavenger receptor cysteine-rich domain homology #status atypical <SRC
F:785-1019/Product: enteropeptidase light chain #status predicted <LCH>
F:785-1014/Domain: trypsin homology <TRY>
F:116-147.179,328,335,388,440,470,503,534,630,682,706,725,848,887,909,949/Binding site:
F:772-896,810-826,910-977,941-956,967-995/Disulfide bonds: #status predicted
F:825,876,971/Active site: His, Asp, Ser #status predicted

Query Match 24.9%; Score 582.5; DB 1; Length 1019;
Best Local Similarity 33.8%; Pred. No. 3.2e-40;
Matches 135; Conservative 62; Mismatches 132; Indels 71; Gaps 15;

QY 67 HF-----IPKQLCDGELDCPLGDEEHCVKSFPGPAVAVRLSKDRSLQVLDSATG 119
Db 647 HFCKNGECVPLNLCDHGLHCEGSDGDEACVRF-----NGTTN 686
QY 120 N-----WFSACFDNFTALSTACRQWY-----SSKPTFRAVEIGPPDDLDVVE 164
Db 687 NGLIVAFRIQSIWHTACAEWTTQISNDVQLGLGSGNSKRFIF-STDGPPFVKLTAP 745

```

```

QY 165 -----ITENSQELMRNNSGPGCLSGSLVSLHC--LACGSLK-----TPRVVGEEASVDSW 214
Db 746 DGHILTPSQ-----CLQDSLIRLQCNHKSCKGKLAQDITPKIVGSGNAKEGAW 796
QY 215 PWQVSIQDKQHVCGGSLDHPHVLTAHFC-PRKHTDVFNNKVRAGSDKLSGSPSLAVAK 273
Db 797 PMVGLYXGRLLCGASLVSSDWLSAAHCVVGRNLEPSKWTAILGLHMKSLTSPQTPV 856
QY 274 III--IEFNPMY---PKDNIALMKLOPPLTFSGVTRVTCICLPFFDEELTPATPLMIWG 328
Db 857 RLIDIVINPHNRRKNDIAMHLEFKVNYTDVIQICLPEENQVPPGPGNCISAGWG 916
QY 329 FTKQNGKMSDILLOASVQVIDSTFCNADDDAYQGVVTEKMKCAGIPGEGVDTCCQDSGGP 388
Db 917 -TWVYQGTANILQADVPELLSNERCQ-QOMPEYNITENNICAGYEKGIDSCQDSGGP 974
QY 389 LMYQ-SDQWHVVGVISWGVCGGPGSTPGVYTKVSAYLNWI 427
Db 975 LMCQENRWFUAGVTSFGYKCALPNRPGVIARVSRFTWI 1014

RESULT 3
A43090
enteropeptidase (EC 3.4.21.9) precursor [validated] - bovine
N:Alternate names: enterokinase
C:Species: Bos primigenius taurus (cattle)
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 28-Apr-2003
C:Accession: A43090; A48874; A61436
R:Kitamoto, Y.; Yuan, X.; Wu, Q.; McCourt, D.W.; Sadler, J.E.
Proc. Natl. Acad. Sci. U.S.A. 91, 7588-7592, 1994
A:Title: Enterokinase, the initiator of intestinal digestion, is a mosaic protease composed
A:Reference number: A43090; MUID:94329561; PMID:8052624
A:Accession: A43090
A:Status: nucleic acid sequence not shown; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1035 <KIT>
A:Cross-references: GB:U09859; NID:g746410; PIDN:AAB40026.1; PID:g746411
A:Experimental source: small intestine
R:LaVallie, E.R.; Rehentulla, A.; Racie, L.A.; DiBlasio, E.A.; Ferenz, K.L.; I
J. Biol. Chem. 268, 23311-23317, 1993
A:Title: Cloning and functional expression of a cDNA encoding the catalytic subunit of bo
A:Reference number: A48874; MUID:94043122; PMID:8226855
A:Accession: A48874
A:Molecule type: mRNA
A:Residues: 801-1035 <LAV>
A:Cross-references: GB:L19663; NID:g416131; PIDN:AAAL6035.1; PID:g416132
A:Note: Parts of this sequence, including the amino end of the mature protein, were confi
R:Light, A.; Janska, H.
J. Protein Chem. 10, 475-480, 1991
A:Title: The amino-terminal sequence of the catalytic subunit of bovine enterokinase.
A:Reference number: A61436; MUID:92189715; PMID:1799406
A:Accession: A61436
A:Molecule type: protein
A:Residues: 801-807, 'Y', 809-827 <LIG>
C:Comment: The mechanism of association with the membrane of the intestinal brush border
embrane attachment using a signal-anchor sequence.
C:Comment: Conversion from membrane-bound to soluble forms may involve further processing
C:Complex: mature enteropeptidase is variously reported to contain two (heavy and light)
liffe linked
C:Function:
A:Description: cleaves propeptide from trypsinogen to produce active trypsin
A:Pathway: intestinal digestive hydrolase cascade
C:Superfamily: enteropeptidase; C1r/C1s repeat homology; LDL receptor ligand-binding rep
C:Keywords: glycoprotein; hydrolase; intestine; serine proteinase; transmembrane protein;
F:22-38/Domain: transmembrane #status predicted <TM>
F:52-117/Product: enteropeptidase mini chain #status predicted <MCH>
F:118-800/Product: enteropeptidase heavy chain #status predicted <HCH>
F:199-236/Domain: LDL receptor ligand-binding repeat homology <LDL1>
F:358-520/Domain: NAM homology <NAM>
F:542-647/Domain: C1r/C1s repeat homology <C1R>
F:659-693/Domain: LDL receptor ligand-binding repeat homology <LDL2>
F:694-799/Domain: scavenger receptor cysteine-rich domain homology #status atypical <SRCF
F:801-1035/Product: enteropeptidase light chain #status predicted <LCH>

```

1;801-1030/Domain: trypsin homology <TRY>
Query Match 24.4%; Score 571; DB 1; Length 1035;
Best Local Similarity 35.0%; Pred. No. 2.9e-39;
Matches 133; Conservative 67; Mismatches 138; Indels 42; Gaps 15;
69 IPKQLCDGELDCPLGEDEHCHVKSPFPGPAVAVRLSKDRSTLOVLDSATGNWFSACPDN 128
672 IPLVNLCDGPFCHKQDSDEAHCVRLF-NGTDSGLVQFR--IQSI-----WHVACAE 722
129 FTEALAEACQMGY-----SKPTTFAVEIGPDQDLVVE-----ITENSQELMRNNSGP 180
723 WTTQISDVDCOLLGLGTGNSVPTF-STGGGYPVNLNTAPNGSLILTPSQ----- 772
181 CLSGSLVSLHC--LACGSLKT----PRVVGGEASVDSWPKQVSIQYDKOHVCGSGTLD 234
773 CLESLILLQCNKYSCGKKLVTOEVSPIKIVGSDSREGAPWVVALYFDQDQVCGASLVS 832
235 PHWVLTAAHC-FRKHTDVENKVRAGSKLSGF--PSLAVAKIIIEFNPMY---PKDND 288
833 RDLVLSAAHCYVGRNNEPSKWKAVLGLHMASNLTSQIETRLIDQIVNPHYNKERRKND 892
289 IALMKLOPLTFSGVTRICLPFPDEEHTPATPLWIGWFTKQNGKMSDILLOASTVQV 348
893 IAMHLEKNAVYDIQIICLPFEENQVPPGRCISAGWALIQ-GSTADVLQADVPL 951
349 IDSTRCNADDAVQGVETKMKCAGIPGGVDTCQDGGGGLMYQ--SDQWHVVGVIS 407
952 LSNKCKQ-QQPEYNIENNVKAGVEAGVDSQDGGGLMCMQENNRWLLAGVTSFGY 1010
408 CGGSPSTPGVYTKVSAYLWNI 427
1011 CALENRPGVYVRPTEWI 1030
RESULT 4
A:Species: Rattus norvegicus (Norway rat)
A:Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 18-Jun-1999
A:Accession: S33777; S32013
A:Title: Cloning and sequence analysis of rat hepsin, a cell surface serine proteinase.
A:Reference number: S33777; MUID:93305733; PMID:8318546
A:Accession: S33777
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-416 <FAR>
A:Cross-references: EMBL:X70900; NID:G57928; PIDN:CAA50256.1; PID:G57929
A:Keywords: hepsin; trypsin homology
A:Superfamily: hepsin; trypsin homology
A:Keywords: hydrolase; liver; serine proteinase; transmembrane protein
F:22-44/Domain: transmembrane #status predicted <TM>
F:162-399/Domain: trypsin homology <TRY>
F:187-203,290-358,321-337,348-380/Disulfide bonds: #status predicted
F:202,256,352/Active site: His, Asp, Ser #status predicted
Query Match 24.3%; Score 568.5; DB 1; Length 416;
Best Local Similarity 30.5%; Pred. No. 1.5e-39;
Matches 137; Conservative 64; Mismatches 155; Indels 93; Gaps 11;
19 KPRIPMETFRKVGIPPIIALLSLASIIIVLVILKYLKQPLHFFPRKQLCDGE 78
14 RPKVAALT---VGLTFITGGAASWALTILR----- 44
79 LDCPLGEDEHCHVKSPFPGPAVAVRLSKDRSTLOVLDSATGNWFSACPDNFEALAEATAC 138
45 -----SQE-----PLYQVQLSPQDSRLVLDKTEGTWRLLCSSRSNARVAGLGC 89
139 RQMGY-----SSKPTTFAVEIG-----PDQLDVVEITENSQELMRNNS 178

Db 90 EEMFLRALAHSELDTVTAGANTSGFFCVDEGGLPLAQRLLDVITSCD----- 138
QY 179 GPCLSGSLVSLHCLACG-KSLKTRPVVVGGEASVDSWPKQVSIQYDKOHVCGSGTLDPEHW 237
Db 139 --CPRGFLPATCQDCGRKLPVDRIVGGQSSIGRWPKVQSLRYDGTGLCGSLSDGW 196
QY 238 VLTAACHFRKHTDYFN-WKVRAGSKDLGSPSLAVAKIIIEFNPMY-----KND 288
Db 197 VLTAACHFRKHTDYFN-WKVRAGSKDLGSPSLAVAKIIIEFNPMY-----KND 288
QY 289 IALMKLOPLTFSGVTRICLPFPDEEHTPATPLWIGWFTKQNGKMSDILLOASTVQV 348
Db 257 IALVHLSSPLTETIIOFVCLPAAGALVDKQVTVTWGNT-QFYGOAVVLOEARVPI 315
QY 349 IDSTRCNADDAVQGVETKMKCAGIPGGVDTCQDGGGGLMYQ--SDQWHVVGVIS 403
Db 316 ISNEVCKSPFYGNQIKFKMFCAGYPEGIDACQDGGGHFVCDRISGTSRWRLCGIVS 375
QY 404 WGYCGGSPSTPGVYTKVSAYLWNIYNWK 432
Db 376 WGTGCLARKPGVYTKVIDFREWIFQAIK 404
RESULT 5
A:Species: Sus scrofa domestica (domestic pig)
A:Date: 07-Oct-1994 #sequence_revision 09-Aug-1996 #text_change 28-Apr-2003
A:Accession: A53663
A:Title: Structural characterization of porcine enteropeptidase.
A:Reference number: A53663; MUID:94327548; PMID:8051081
A:Accession: A53663
A:Molecule type: mRNA
A:Residues: 1-1034 <NAT>
A:Cross-references: GB:D30799; NID:G505122; PIDN:BA06459.1; PID:G505123
A:Note: parts of this sequence, including the amino ends of three chains isolated from the
C:Comment: The mechanism of association with the membrane of the intestinal brush border
C:Complex: Mature enteropeptidase is variously reported to contain two (heavy and light)
ed by a disulfide bond. Possibly, conversion from membrane-bound to soluble forms involve
C:Function:
A:Description: cleaves activation peptide from trypsinogen to produce active trypsin
A:Pathway: intestinal digestive hydrolase cascade
C:Superfamily: enteropeptidase; Ctr/Cis repeat homology; LDL receptor ligand-binding rep
C:Keywords: glycoprotein; hydrolase; serine proteinase; transmembrane protein; zymogen
F:22-38/Domain: transmembrane #status predicted <TM>
F:52-117/Product: enteropeptidase mini chain #status predicted <MCH>
F:118-799/Product: enteropeptidase heavy chain #status predicted <HCH>
F:199-236/Domain: LDL receptor ligand-binding repeat homology <LDL1>
F:357-519/Domain: MAM homology <MAM>
F:541-646/Domain: Ctr/Cis repeat homology <CLR>
F:658-692/Domain: LDL receptor ligand-binding repeat homology <LDL2>
F:693-798/Domain: scavenger receptor cysteine-rich domain homology #status atypical <SRCF
F:800-1034/Product: enteropeptidase light chain #status predicted <LCH>
F:116,147,170,194,283,343,350,403,455,485,518,549,697,701,721,740,761,804,863,902,964
F:787-911,825-841,925-992,956-971,982-1010/Disulfide bonds: #status predicted
F:840,986/Active site: His, Asp, Ser #status predicted
Query Match 23.6%; Score 552; DB 1; Length 1034;
Best Local Similarity 34.2%; Pred. No. 1.1e-37;
Matches 127; Conservative 69; Mismatches 141; Indels 34; Gaps 14;
QY 74 LDCPLGEDEHCHVKSPFPGPAVAVRLSKDRSTLOVLDSATGNWFSACPDNFEAL 133
Db 676 LDCGFSHCKDGSDEAHCVR-FLNGTANNSGLVQFR--IQSI-----WHVACAEWITQT 726
QY 134 AETHACRQMGY-----SSKPTTFAVEIGPDQLDVVEITENSQELMRNNSGCLSGSLVSL 189

F;20-390/Product: plasma kallikrein heavy chain #status experimental <MAT1>
F;20-109/Domain: apple repeat <AP1>
F;110-199/Domain: apple repeat <AP2>
F;200-289/Domain: apple repeat <AP3>
F;291-380/Domain: apple repeat <AP4>
F;391-638/Product: plasma kallikrein light chain #status experimental <MAT2>
F;391-621/Domain: trypsin homology <TRY>
F;21-104,47-77,51-57,111-194,137-166,141-147,201-284,227-256,231-237,292-375,318-347,322-324,325-328,330-332,333-335,336-338,339-341,342-344,345-347,348-350,351-353,354-356,357-359,360-362,363-365,366-368,369-371,372-374,375-377,378-380,381-383,384-386,387-389,390-392,393-395,396-398,399-401,402-404,405-407,408-410,411-413,414-416,417-419,420-422,423-425,426-428,429-431,432-434,435-437,438-440,441-443,444-446,447-449,450-452,453-455,456-458,459-461,462-464,465-467,468-470,471-473,474-476,477-479,480-482,483-485,486-488,489-491,492-494,495-497,498-500,501-503,504-506,507-509,510-512,513-515,516-518,519-521,522-524,525-527,528-530,531-533,534-536,537-539,540-542,543-545,546-548,549-551,552-554,555-557,558-560,561-563,564-566,567-569,570-572,573-575,576-578,579-581,582-584,585-587,588-590,591-593,594-596,597-599,600-602,603-605,606-608,609-611,612-614,615-617,618-620,621-623,624-626,627-629,630-632,633-635,636-638,639-641,642-644,645-647,648-650,651-653,654-656,657-659,660-662,663-665,666-668,669-671,672-674,675-677,678-680,681-683,684-686,687-689,690-692,693-695,696-698,699-701,702-704,705-707,708-710,711-713,714-716,717-719,720-722,723-725,726-728,729-731,732-734,735-737,738-740,741-743,744-746,747-749,750-752,753-755,756-758,759-761,762-764,765-767,768-770,771-773,774-776,777-779,780-782,783-785,786-788,789-791,792-794,795-797,798-800,801-803,804-806,807-809,810-812,813-815,816-818,819-821,822-824,825-827,828-830,831-833,834-836,837-839,840-842,843-845,846-848,849-851,852-854,855-857,858-860,861-863,864-866,867-869,870-872,873-875,876-878,879-881,882-884,885-887,888-890,891-893,894-896,897-899,900-902,903-905,906-908,909-911,912-914,915-917,918-920,921-923,924-926,927-929,930-932,933-935,936-938,939-941,942-944,945-947,948-950,951-953,954-956,957-959,960-962,963-965,966-968,969-971,972-974,975-977,978-980,981-983,984-986,987-989,990-992,993-995,996-998,999-1001,1002-1004,1005-1007,1008-1010,1011-1013,1014-1016,1017-1019,1020-1022,1023-1025,1026-1028,1029-1031,1032-1034,1035-1037,1038-1040,1041-1043,1044-1046,1047-1049,1050-1052,1053-1055,1056-1058,1059-1061,1062-1064,1065-1067,1068-1070,1071-1073,1074-1076,1077-1079,1080-1082,1083-1085,1086-1088,1089-1091,1092-1094,1095-1097,1098-1100,1101-1103,1104-1106,1107-1109,1110-1112,1113-1115,1116-1118,1119-1121,1122-1124,1125-1127,1128-1130,1131-1133,1134-1136,1137-1139,1140-1142,1143-1145,1146-1148,1149-1151,1152-1154,1155-1157,1158-1160,1161-1163,1164-1166,1167-1169,1170-1172,1173-1175,1176-1178,1179-1181,1182-1184,1185-1187,1188-1190,1191-1193,1194-1196,1197-1199,1200-1202,1203-1205,1206-1208,1209-1211,1212-1214,1215-1217,1218-1220,1221-1223,1224-1226,1227-1229,1230-1232,1233-1235,1236-1238,1239-1241,1242-1244,1245-1247,1248-1250,1251-1253,1254-1256,1257-1259,1260-1262,1263-1265,1266-1268,1269-1271,1272-1274,1275-1277,1278-1280,1281-1283,1284-1286,1287-1289,1290-1292,1293-1295,1296-1298,1299-1301,1302-1304,1305-1307,1308-1310,1311-1313,1314-1316,1317-1319,1320-1322,1323-1325,1326-1328,1329-1331,1332-1334,1335-1337,1338-1340,1341-1343,1344-1346,1347-1349,1350-1352,1353-1355,1356-1358,1359-1361,1362-1364,1365-1367,1368-1370,1371-1373,1374-1376,1377-1379,1380-1382,1383-1385,1386-1388,1389-1391,1392-1394,1395-1397,1398-1399,1400-1402,1403-1405,1406-1408,1409-1411,1412-1414,1415-1417,1418-1420,1421-1423,1424-1426,1427-1429,1430-1432,1433-1435,1436-1438,1439-1441,1442-1444,1445-1447,1448-1450,1451-1453,1454-1456,1457-1459,1460-1462,1463-1465,1466-1468,1469-1471,1472-1474,1475-1477,1478-1480,1481-1483,1484-1486,1487-1489,1490-1492,1493-1495,1496-1498,1499-1501,1502-1504,1505-1507,1508-1510,1511-1513,1514-1516,1517-1519,1520-1522,1523-1525,1526-1528,1529-1531,1532-1534,1535-1537,1538-1540,1541-1543,1544-1546,1547-1549,1550-1552,1553-1555,1556-1558,1559-1561,1562-1564,1565-1567,1568-1570,1571-1573,1574-1576,1577-1579,1580-1582,1583-1585,1586-1588,1589-1591,1592-1594,1595-1597,1598-1599,1600-1602,1603-1605,1606-160

F;21-104,47-77,51-57,111-194,137-166,141-147,201-284,227-256,231-237,292-375,318-347,322
 F;127,215,308,396,494/Binding site: carbohydrate (asn) (covalent) #status predicted
 F;434,483,578/Active site: His, Asp, Ser #status predicted

Query Match 21.7%; Score 508; DB 1; Length 638;
 Best Local Similarity 41.0%; Pred. No. 2.9e-34;
 Matches 112; Conservative 45; Mismatches 82; Indels 34; Gaps 10;

QY 174 MRNSGPGCLSGSLVSLHCLACGSLKTP-----RVVGGEEASVDSWPMQVSIQ---Y 222
 DB 364 MQGSSG-----YSRLCKLVDSPTCTTKINARIIVGGTINASLGEWPMQVSLQKLV 413
 QY 223 DKQHVCGGSLDPHWLTAHCFER--KHTDVFNWKVRAG-----SDKLGSPSPSLAVAKII 276
 DB 414 SQTHLCGGSLIGQWVLTAAHCFDGPDPDV--WRIYGGILSLSEITKTPSRIKELII 471
 QY 277 IEFNPMYKNDIALMKLOPFTFSQVTRPICLPFFDEBELTATPLWIIIGWFTKQNGGK 336
 DB 472 HQEYKVSQGNVDIALKLOPLNNTYEFQKPICLPSKADNTIIVTNCWVTGWYTKEQ-GE 530
 QY 337 MSDILLOASVQVDSIDTRCNADDAVQGEVTEK-MWCAGIPEGGVDTCQDGGGGLMYQ-SD 394
 DB 531 TQNILOKATIPLVNBECC--QKKYRDYVINKQMICAGYKEGGTDACKDGGGGLVCKHSG 588
 QY 395 QWHVVGIVSWGCGGPGSPGVYTKVSAYLWMI 427
 DB 589 RWQLVGITWSGCGGKQDQPGVYTKVSEWDMI 621

RESULT 8

JX0172
 A:acrosin (EC 3.4.21.10) precursor form 3 - mouse
 C:Species: Mus musculus (house mouse)
 C>Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 20-Jun-2000
 C:Accession: JX0172; JX0138
 J. J. Biochem. 109, 828-833, 1991
 A:Title: Structure and organization of the mouse acrosin gene.
 A:Reference number: JX0172; MUID:92041732; PMID:1939002
 A:Accession: JX0172

A:Molecule type: DNA
 A:Residues: 1-436 <WAT>
 A:Cross-references: GB:S66245; NID:9238706; PIDN:AAB20293.1; PID:9238707
 A:Title: Kashiwabara, S.; Baba, T.; Takada, M.; Watanabe, K.; Yano, Y.; Arai, Y.
 J. Biochem. 108, 785-791, 1990
 A:Title: Primary structure of mouse proacrosin deduced from the cDNA sequence and its gene
 A:Reference number: JX0138; MUID:91185335; PMID:2127931
 A:Accession: JX0138
 A:Molecule type: mRNA
 A:Residues: 4-436 <KAS>
 A:Cross-references: GB:D00754; NID:9220322; PIDN:BA00651.1; PID:9220323
 C:Comment: Acrosin is an acrosomal protease that plays an important role in the initial
 e-binding activity.

C:Genetics:
 A:Introns: 26/2; 95/2; 190/1; 238/3
 C:Superfamily: acrosin; trypsin homology
 C:Keywords: glycoprotein; hydrolase; serine proteinase; sperm
 F;1-19/Domain: signal sequence #status predicted <SIG>
 F;20-42/Product: acrosin light chain #status predicted <ALC>
 F;43-321/Product: acrosin heavy chain #status predicted <AHC>
 F;43-286/Domain: trypsin homology <TRY>
 F;22,211/Binding site: carbohydrate (Asn) (covalent) #status predicted
 F;25-155,29-162,74-90,178-247,210-226,237-267/Disulfide bonds: #status predicted
 F;89,143,241/Active site: His, Asp, Ser #status predicted
 F;7235/Binding site: substrate (Asp) #status predicted

Query Match 21.6%; Score 505; DB 2; Length 436;
 Best Local Similarity 40.9%; Pred. No. 3.2e-34;
 Matches 106; Conservative 37; Mismatches 90; Indels 26; Gaps 7;

QY 194 CG-----KSLKTPRVVGGEEASVDSWPMQVSIQY-----DKQHVCGGSLDPHWLTA 242
 DB 29 CGLRFRNSQAGTRIVSGSLQGLGAWPMWVSIQITSHNSRYHACGSLNSHWLTA 88

QY 243 HCFRKHDTDFVNWK-----VRAGSKLGSPFSLA--VAKIIIEFNPMYKNDIALMK 293
 DB 89 HCFDNKKVYDWRVLVFGAQEIYGVGRNKPVEQQRVYQKIVHEKYNVVTGNDIALLK 148
 QY 294 LQFPITFSGTVRPICLPFPDEBLTATP--LWIIIGWFTKQNGGKMSDILLOASVQVDS 351
 DB 149 ITPPVTCGNFIGPCCLPHF-KAGPPQIPIHTCVVTGWGIKEKAPRSPVIMEARVDLID 207
 QY 352 TRCNADDAVQGEVTEKMKCAGIPEGGVDTCQDGGGGLMYQSD---QWHVVGIVSWGVC 408
 DB 208 DUCNSTQWNGRVTSTNVCAGYPEGKIDTCQDGGGGLMCRDNDVSPFVVVGITSMGVC 267
 QY 409 GGPSTPGVYTKVSAYLWMI 427
 DB 268 ARAKRPGVYATWDYLDWI 286

RESULT 9

T30337
 A:polyprotein - African clawed frog
 C:Species: Xenopus laevis (African clawed frog)
 C>Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 03-Feb-2003
 C:Accession: T30337
 R. Yang, J.C.; Lindsay, L.B.; Hedrick, J.L.
 submitted to the EMBL Data Library, March 1998
 A:Description: cDNA cloning of ovocytinase, a chymotrypsin-like protease released from Xer
 A:Reference number: Z20829
 A:Accession: T30337
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-1524 <YAN>
 A:Cross-references: EMBL:U81290; NID:92981640; PID:92981641; PIDN:AAC24717.1
 C:Superfamily: trypsin related polyprotein; trypsin homology

Query Match 21.5%; Score 504.5; DB 2; Length 1524;
 Best Local Similarity 44.0%; Pred. No. 1.6e-33;
 Matches 109; Conservative 38; Mismatches 86; Indels 15; Gaps 8;

QY 194 CGKSLKT-----PRVVGEEASVDSWPMQVSIQYDKQHVCGGSLDPHWLTAACFRKH 248
 DB 570 CGWAPMTKWLPRIVGGEASPSWPMQVIFLRTTHCEGAIISQWILTAAHCIRRA 629
 QY 249 TDVFNWKRAGS-DKLGSPSLAVAKIIIEFNPMYK---DNIDALMKLOPPLTFSGTV 304
 DB 630 EFSY-WTVIAGDHRMLNLESTEQIRNIKTIRHDYNSETVDNDIALYLRPLDNDV 688
 QY 305 RPICLPFPDEBLTATPLWIIIGWFTKQNGGKMSDILLOASVQVDSIDTRCNADDAVQGEV 364
 DB 689 RPVCLPPEEVLTPASVCVVTGWGNTAEDGQFALG-LQQLQLPILDSITCNT-SYISGEL 746
 QY 365 TEKMWKAGIPEG-GVDTCCQDGGGGLMYQS--DQWHVVGIVSWGCGGSPTPGVYTKVS 421
 DB 747 TDHMLCAGFPSSKEKACQDGGGGLVQCNEKEQFSIYGLVSWGEGGRVSKPGVYTKV 806
 QY 422 AYLNWYIN 429
 DB 807 LFFTWION 814

RESULT 10

A37344
 A:acrosin (EC 3.4.21.10) precursor form 1 - mouse
 C:Species: Mus musculus (house mouse)
 C>Date: 19-Mar-1993 #sequence_revision 19-Mar-1993 #text_change 22-Jun-1999
 C:Accession: A37344
 R. Klemm, U.; Maier, W.M.; Tsoulosidou, S.; Adham, I.M.; Willison, K.; Engel, W.
 Differentiation 42, 160-166, 1990
 A:Title: Mouse proacrosin: cDNA sequence, primary structure and postmeiotic expression
 A:Reference number: A37344; MUID:90255839; PMID:2111255
 A:Accession: A37344
 A:Status: preliminary
 A:Molecule type: mRNA

A:Residues: 1-418 <KLE>
A:Cross-references: GB:X52466; NID:G49857; PIDN:CAA36704.1; PID:G49858
C:Superfamily: acrosin; trypsin homology
C:Keywords: glycoprotein; hydrolase; serine proteinase
F:39-283/Domain: trypsin homology <Try>
F:16-208/Binding site: carbohydrate (asn) (covalent) #status predicted
F:21-151/Disulfide bonds: #status predicted
F:25-158/Disulfide bonds: #status predicted
F:70-86/Disulfide bonds: #status predicted
F:185,139,238/Active site: His, Asp, Ser #status predicted
F:175-244/Disulfide bonds: #status predicted
F:207-223/Disulfide bonds: #status predicted
F:234-264/Disulfide bonds: #status predicted

Query Match 21.5%; Score 503; DB 2; Length 418;
Best Local Similarity 39.1%; Pred. No. 4.4e-34;
Matches 108; Conservative 36; Mismatches 90; Indels 42; Gaps 7;

QY	179	GPCLSGLSVLHCLACKSKLT-----PRVVGEEASVDSWPMQVSIQY-----DKQ	225
DB	23	GPC-----GLRTRQSQAGTRIVSGOSAHVGAWPMVSLQIFTHSNRRY	67
QY	226	HVCGGSGILDPHWLVTAAHCFRKHTDVFNMK-----VRAGSDKLSGFP--SLAVAKIII	276
DB	68	HACGGSLLNSHWLVTAAHCFDNKKVYDMLVFGAQEIYGRNRPVKPEERYVQKIVI	127
QY	277	IEFNPMYPKDNIDIALMKLQPLFTFSGTVRPICLPFFDEELTRATP--LWTIGWGFTKQNG	334
DB	128	HEKYNVVTGNDIALLKVTTPVPCGNFIGFCCLPHFKAGPPRKIPHTCYVTGHWYIKREA	187
QY	335	GKMSDILLQASVOVISTRCNADDAVQGETEKKMCAGIPEGGVDTCCQSDSGGLPMYQSD	394
DB	188	PRPSVYMEARVDLIDLCLNSTQWINGRVTSTNVCAGYPEGKIDICQSDSGGLMCRDN	247
QY	395	---QHWVGVISVNGYCGGSGSTGVYTKVSAYLNWI	427
DB	248	ARQPFVVGILTSNGVGCARAKREGVTTATWDYLDWI	283

RESULT 11
S18407
acrosin (EC 3.4.21.10) precursor - rat
N1Contains: proacrosin
C1Species: Rattus norvegicus (Norway rat)
C1Date: 31-Dec-1993 #sequence revision 31-Dec-1993 #text_change 22-Jun-1999
C1Accession: S18407; S30037; A56620
R1K1Klemm, U.; Flake, A.; Engel, W.
Biochim. Biophys. Acta 1090, 270-272, 1991
A1Title: Rat sperm acrosin: cDNA sequence, derived primary structure and phylogenetic or
A1Reference number: S18407; MUID:92031708; PMID:1932123
A1Accession: S18407
A1Molecule type: mRNA
A1Residues: 1-437 <KLE>
R1K1Klemm, U.; Flake, A.; Engel, W.
submitted to the EMBL Data Library, April 1991
A1Reference number: S30037
A1Accession: S30037
A1Molecule type: mRNA
A1Residues: 1-254, 'LCDR', 259, 'DHFL', 264, 'GRLC', 269-437 <KLE2>
A1Cross-references: EMBL:X59254; NID:G57282; PIDN:CAA41947.1; PID:G57283
R1K1Krenning, H.; Flake, A.; Adham, I.M.; Radtke, J.; Engel, W.
DNA Seq. 2, 57-60, 1991
A1Title: Exon-intron structure and nucleotide sequence of the rat proacrosin gene.
A1Reference number: A56620; MUID:92199245; PMID:1802037
A1Accession: A56620
A1Status: preliminary
A1Molecule type: DNA; mRNA
A1Residues: 1-254, 'LCDR', 259, 'DHFL', 264, 'GRLC', 269-437 <KRE>
A1Note: sequence modified after extraction from NCBI backbone
A1Note: sequence extracted from NCBI backbone (NCBIN:89436, NCBIN:89439, NCBIN:89447, NC
C1Superfamily: acrosin; trypsin homology
C1Keywords: glycoprotein; hydrolase; serine proteinase; sperm; zymogen

Db 189 PRSPVLLARVELIDLCLNCTWNGRWSTNCAGYPEGKIDTCQDPSGGPLMCRON 248
QY 393 -SDQHWVGVTSWGYCGGSPSTPGVTKVSAYLNWI 427
Db 249 ANSPFVVVGITSGWGCARAKREGIYATWDYLDWI 284

RESULT 13
KOHUP
plasma kallikrein (EC 3.4.21.34) precursor - human
N:Alternate names: kininogenin; plasma prekallikrein
C:Species: Homo sapiens (man)
C:Date: 13-Aug-1986 #sequence_revision 13-Aug-1986 #text_change 18-Jun-1999
C:Accession: A00921; A37939
R:Chung, D.W.; Fujikawa, K.; McMullen, B.A.; Davie, E.W.
Biochemistry 25, 2410-2417, 1986
A:Title: Human plasma prekallikrein, a zymogen to a serine protease that contains four
A:Reference number: A00921; MUID:86243359; PMID:3521732
A:Accession: A00921
A:Molecule type: mRNA
A:Residues: 1-638 <CHU>
A:Cross-references: GB:M13143; NTD:gl90262; PIDN:AAA60153.1; PID:gl90263
R:McMullen, B.A.; Fujikawa, K.; Davie, E.W.
Biochemistry 30, 2050-2056, 1991
A:Title: Location of the disulfide bonds in human plasma prekallikrein: the presence of
A:Reference number: A37939; MUID:91152016; PMID:1998666
A:Accession: A37939
A:Molecule type: protein
A:Residues: 20-27;40-46,'X',48,'H',50,'X',52-70,'H',75-76,'X',78-80;103-113;131-140;141-
142;160-283,'X',285;287-291,'X',293-295;314-317,'X',319-320;321-324,'X',329-333;334-339,'X',
525;538-551;562,'X',564-567;573,'X',575-576;578-583,'X',585;592-604 <MCM>
C:Comment: This protein, synthesized in the liver, circulates as a noncovalent complex w
C:Comment: The zymogen is activated by factor XIIa, which cleaves the molecule into a li
A: linked by one or more disulfide bonds.
C:Comment: the enzyme cleaves Lys-Arg and Arg-Ser bonds. It activates, in a reciprocal
X: kininogen and may also play a role in the renin-angiotensin system by converting prorenin i
A: C:Genetics:
A:Gene: GDB:KLK3
A:Cross-references: GDB:127575; OMIM:229000
A:Map position: 4q35-q35
C:Superfamily: coagulation factor XI; trypsin homology
C:Keywords: blood coagulation; duplication; fibrinolysis; glycoprotein; hydrolase; infla
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-638/Product: plasma kallikrein #status predicted <NAT>
F:390-390/Domain: plasma kallikrein heavy chain #status predicted <HCH>
F:20-109/Domain: apple repeat <AP1>
F:110-195/Domain: apple repeat <AP2>
F:200-289/Domain: apple repeat <AP3>
F:291-380/Domain: apple repeat <AP4>
F:391-621/Domain: trypsin homology <TRY>
F:391-638/Domain: plasma kallikrein light chain #status predicted <LCH>
F:21-104,47-77,51-57,111-194,137-166,141-147,201-284,227-256,231-237,292-375,322-328,383
F:127,308,396,453,494/Binding site: carbohydrate (Asn) (covalent) #status experimental
F:138-347,340-345/Disulfide bonds: #status predicted
F:390-391/Cleavage site: Arg-Ile (coagulation factor XIIa) #status predicted
F:434,483,578/Active site: His, Asp, Ser #status predicted

Query Match 21.0%; Score 491.5; DB 1; Length 638;
Best Local Similarity 42.2%; Pred. No. 6.8e-33;
Matches 100; Conservative 47; Mismatches 74; Indels 16; Gaps 8;

QY 202 RVVGGEEASVDSWPQVSIQYD--KQHVCGSILDPHWLTAHCF--RKHTDYFNKVV 256
Db 390 RVVGTNSWGEWPNQVSLQVLTQRLCGSLGHQWLVTAHCFDGLPLQDY--WRI 447

QY 257 PAG-----SDKLGSPFSLAVAKIIITFEFPMYKNDIALMKLPPLTFSGTVRPICLPFF 312
Db 448 YSGLNLSDITKDTFSPQIKELIIHQNVKSEGNHDIKLOAPLNTFQKPICLPSK 507

QY 313 DEELTPATPLMIIGFTKQNGKNSDILQASVQVVIDSTRCNADDDAYQG-EVTEKMWCA 371
Db 508 GDTSTIYNCWTVGTWGFSEK-GEIQLNQVNIPLVNEEC--QKRQDYKITQRWCA 564

QY 372 GIPEGVDTCQDSSGGLMYQ--SDQHWVGVTSWGYCGGSPSTPGVTKVSAYLNWI 427
Db 565 GYKEGGKDAKGDGSGPLCKYKNGWRLVGLTSGEGCARRQPGVYTKVAETMDWI 621

RESULT 14
JE0315
low-density lipoprotein receptor-related protein - mouse
C:Species: Mus musculus (house mouse)
C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 03-Feb-2003
C:Accession: JE0315
R:Tomita, Y.; Kim, D.; Megoori, K.; Fujino, T.; Yamamoto, T.T.
J. Biochem. 124, 784-789, 1998
A:Title: A novel low-density lipoprotein receptor-related protein with type II membrane
A:Reference number: JE0315; MUID:98429596; PMID:9756624
A:Accession: JE0315
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1113 <TOM>
A:Cross-references: DBJ:AB013874; NID:g3869144; PIDN:BAA34371.1; PID:g3869145
C:Superfamily: mouse low-density lipoprotein receptor-related protein; LDL receptor ligand
F:337-372/Domain: LDL receptor ligand-binding repeat homology <LDL1>
F:374-408/Domain: LDL receptor ligand-binding repeat homology <LDL2>
F:410-445/Domain: LDL receptor ligand-binding repeat homology <LDL3>
F:447-482/Domain: LDL receptor ligand-binding repeat homology <LDL4>
F:484-682/Domain: LDL receptor ligand-binding repeat homology <LDL5>
F:684-720/Domain: LDL receptor ligand-binding repeat homology <LDL6>
F:723-757/Domain: LDL receptor ligand-binding repeat homology <LDL7>
F:869-1097/Domain: trypsin homology <TRY>

Query Match 20.9%; Score 488.5; DB 2; Length 1113;
Best Local Similarity 33.2%; Pred. No. 2.4e-32;
Matches 130; Conservative 52; Mismatches 148; Indels 61; Gaps 17;

QY 69 IPRKQLCGELDPLGDEBCHVSKFPPGPAVAVPLSKDRSTLOVLDSATGNWFACPDN 128
Db 736 VPRDLWCDGWDCSDSDSEWCGVTLSKNG-----NSSLLTVHKSAREH--HVCADG 785

QY 129 FTEALAEACRQMGYSKPTFRAVEIGPDQLDVVEITENSQELRM-----RNSG----- 179
Db 786 WRETLSQLAQKQGLGEPVTKLI---PGQ-----EGQWLLEYNWNLSGTLOE 834

QY 180 -----PCLSGSLVSLHCLA--CGK---SLKTPRVVGGEEASVDSWPQVSIQYDKQ--HV 227
Db 835 LLVPHSCPSSEISLLCSKQDCGRPAARMKMKRILGRTSRPGRWQCSLQSEPSGHI 894

QY 228 CGGSILDPHWLTAHCFRKHDTVFNKVRAGSKL---GSFPFLAVAKIIIEFNPMYP 284
Db 895 CGVLIKKVLTVAHCFEGREDADVVKVFGINNLDHPSGFMQTRFVKTLIL--HPRYS 952

QY 285 K---DNIDIALMKLOPPLTFSGTVRPICLPFFDEELTPATPLMIIGWTKQNGKMSDIL 341
Db 953 RAVVDYDISVVELSDDINEITSYVAPVCLPSPVEELEDPTYIITGWG---HMGKWPFL 1009

QY 342 LQASVQVIDSTRCNADDDAY--QGVTEKMWCAIPEGVDTCQDSSGGLMYQ--SDQHW 397
Db 1010 QEGEVRIPLQEC---QSYFDMKTIITNRMICAGYESGTVDSCMGDSGGLVCPERPQQWT 1066

QY 398 VVGTVSRYGCGGSPST-PGVYTKVSAYLNWI 427
Db 1067 LFGITSWGSVCFSKVLGPGVTSNVSFVGWI 1097

RESULT 15
S11674
acrosin (EC 3.4.21.10) precursor - human
C:Species: Homo sapiens (man)
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 16-Jun-2000
C:Accession: S11674; S23499; S12063; A61022; S03330
R:Keime, S.; Adham, I.M.; Engel, W.
Eur. J. Biochem. 190, 195-200, 1990
A:Title: Nucleotide sequence and exon-intron organization of the human proacrosin gene.
A:Reference number: S11674; MUID:90306003; PMID:21114285

A:ACCESSION: S11674
 A:Molecule type: DNA
 A:Residues: 1-421 <KEI>
 A:Cross-references: EMBL:X54017; NID:g35582; PIDN:CRA37964.1; PID:g1216165
 A:Note: the authors translated the codon AGG for residue 64 as Thr and CTG for residue 2
 R:Vazquez-Levin, M.H.; Revencos, J.; Gordon, J.W.
 Eur. J. Biochem. 207, 23-26, 1992
 A:Title: Molecular cloning, sequencing and restriction mapping of the genomic sequence e
 A:Reference number: S23499; MUID:92331659; PMID:1628652
 A:Accession: S23499
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-421 <VAZ>
 A:Cross-references: EMBL:M77378
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1992
 R:Keime, S.
 submitted to the EMBL Data Library, December 1989
 A:Reference number: S12063
 A:Accession: S12063
 A:Molecule type: DNA
 A:Residues: 1-225, 'R', 227-421 <KEI2>
 A:Cross-references: EMBL:X54017
 R:Adham, I.M.; Klemm, U.; Maier, W.M.; Engel, W.
 Hum. Genet. 84, 125-128, 1990
 A:Title: Molecular cloning of human proacrosin cDNA.
 A:Reference number: A61022; MUID:90128988; PMID:2298447
 A:Accession: A61022
 A:Status: not compared with conceptual translation
 A:Molecule type: mRNA
 A:Residues: 1-63, 'T', 65-225, 'V', 227-267, 'R', 269-421 <ADH>
 R:Baba, T.; Watanabe, K.; Kashiwabara, S.I.; Arai, Y.
 FEBS Lett. 244, 296-300, 1989
 A:Title: Primary structure of human proacrosin deduced from its cDNA sequence.
 A:Reference number: S03330; MUID:89153568; PMID:2493394
 A:Accession: S03330
 A:Molecule type: mRNA
 A:Residues: 1-63, 'T', 65-119, 'V', 121-165, 'L', 167-267, 'R', 269-344, 'R', 346-421 <BAB>
 A:Cross-references: EMBL:I00970; NID:g28325; PIDN:CRA68784.1; FID:g28326
 C:Genetics:
 A:Gene: GDB:ACR
 A:Cross-references: GDB:119645; OMIM:102480
 A:Map position: 22q13-22qter
 A:Introns: 26/2; 94/2; 189/2; 237/3
 C:Superfamily: acrosin; trypsin homology
 C:Keywords: Glycoprotein; hydrolase; serine proteinase; sperm
 F:1-19/Domain: signal sequence #status predicted <SIG>
 F:20-421/Product: acrosin #status predicted <MAT>
 F:20-42/Product: acrosin light chain #status predicted <LCH>
 F:43-421/Product: acrosin heavy chain #status predicted <HCH>
 F:43-285/Domain: trypsin homology <TRY>
 F:302-379/Region: proline-rich
 F:22,210/Binding site: carbohydrate (Asn) (covalent) #status predicted
 F:25-154/Disulfide bonds: #status predicted
 F:29-162/Disulfide bonds: #status predicted
 F:73-89/Disulfide bonds: #status predicted
 F:88,142,240/Active site: His, Asp, Ser #status predicted
 F:177-246/Disulfide bonds: #status predicted
 F:209-225/Disulfide bonds: #status predicted
 F:236-266/Disulfide bonds: #status predicted

Query Match 20.5%; Score 481; DB 1; Length 421;
 Best Local Similarity 41.0%; Pred. No. 3e-32;
 Matches 100; Conservative 39; Mismatches 88; Indels 18; Gaps 5;

Db 162 CLPHFKAGLPRGSOSCWAGWGYIEEKAPRESSILMEARVDLIDLCLNSTQWYNGRVQP 221
 QY 367 KVMCAGIPEGGVDTCCQGSGLMYQSDQ---WHVVGIVSWGYGCGSPSTPGVYTKVSAY 423
 Db 222 TNVCAGYFVGKIDTCQGSGLCKDSKESAYVVVGITTSNGVGCGALAKRFGIYTATWPY 281
 QY 424 LNWI 427
 Db 282 LNWI 285

Search completed: November 23, 2003, 07:52:53
 Job time : 30.7208 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 23, 2003, 07:39:05 ; Search time 16.7538 Seconds
(without alignments)
1221.015 Million cell updates/sec

Title: US-09-607-745-2

Perfect score: 2342

Sequence: 1 MDPDSQPLNSLDVPLRKP.....VTKVSAVLNWIYVWKAEL 435

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	2337	99.8	437	1	TMS4_HUMAN
2	687	29.3	453	1	Q8kit0 homo sapien
3	673.5	28.8	492	1	TMS2_HUMAN
4	660.5	28.2	454	1	TMS3_HUMAN
5	648.5	27.7	490	1	TMS2_MOUSE
6	628.5	26.8	455	1	TMS5_MOUSE
7	584.5	25.0	417	1	HEPS_HUMAN
8	582.5	24.9	1019	1	ENTK_HUMAN
9	578	24.7	436	1	HEPS_MOUSE
10	574	24.5	457	1	TMS5_HUMAN
11	571	24.4	1035	1	ENTK_BOVIN
12	568.5	24.3	416	1	HEPS_RAT
13	568	24.3	418	1	HART_HUMAN
14	552.5	23.6	1069	1	ENTK_MOUSE
15	552	23.6	1034	1	ENTK_PIG
16	511	21.8	638	1	KAL_RAT
17	508	21.7	638	1	KAL_MOUSE
18	505	21.6	436	1	ACRO_MOUSE
19	503.5	21.5	324	1	TEST_MOUSE
20	501.5	21.4	437	1	ACRO_RAT
21	498	21.3	875	1	NETR_HUMAN
22	491.5	21.0	638	1	KAL_HUMAN
23	488.5	20.9	1113	1	COR1_MOUSE
24	484	20.7	1042	1	COR1_HUMAN
25	483	20.6	321	1	TRYG_HUMAN
26	482	20.6	855	1	ST14_MOUSE
27	481	20.5	421	1	ACRO_HUMAN
28	478.5	20.4	415	1	ACRO_PIG
29	478	20.4	314	1	TEST_HUMAN
30	476.5	20.3	311	1	TRYG_MOUSE
31	475	20.3	422	1	DES1_HUMAN
32	472.5	20.2	317	1	BSS4_HUMAN
33	470	20.1	761	1	NETR_MOUSE

34 468.5 20.0 431 1 ACRO_RABIT
35 467.5 20.0 290 1 MPN_HUMAN
36 463 19.8 855 1 ST14_HUMAN
37 461.5 19.7 277 1 TRY2_ANOGA
38 460.5 19.7 263 1 CTBB_RAT
39 459.5 19.6 342 1 PSS8_RAT
40 459.5 19.6 812 1 PLMN_MOUSE
41 458.5 19.6 267 1 TRY7_ANOGA
42 458.5 19.6 790 1 PLMN_PIG
43 457.5 19.5 625 1 FALL_HUMAN
44 452.5 19.3 342 1 PSS8_MOUSE
45 449.5 19.2 333 1 PLMN_CANFA

ALIGNMENTS

RESULT 1
TMS4_HUMAN
ID TMS4_HUMAN STANDARD; PRT; 437 AA.
AC Q9NRS4; Q9NZAS;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Transmembrane protease, serine 4 (EC 3.4.21.-) (Membrane-type serine
DE protease 2) (MT-SP2).
GN TMRSS4 OR TMRSS3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Pancratic carcinoma;
RX MEDLINE=20283276; PubMed=10825129;
RA Wallrapp C., Haehnel S., Mueller-Pillasch F., Burghardt B.,
RA Iwamura T., Ruthenberger M., Lerch M.M., Adler G., Gress T.M.;
RT "A novel transmembrane serine protease (TMRSS3) overexpressed in
RT pancreatic cancer.";
RL Cancer Res. 60:2602-2606(2000).
[2]
RP SEQUENCE FROM N.A.
RA Smeekens S.S., Iorimer D.D., Wang E., Hou J., Linnevers C.;
RT "MT-SP2, a novel type II membrane serine protease expressed in
RT trachea, colon, and small intestine: identification, cloning, and
RT chromosomal localization.";
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
[3]
RP SEQUENCE FROM N.A.
RC TISSUE=Pancreas;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Bustow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh P.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Carninci P., Prange C.,
RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Mullay S.J.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Rulyk S.W.,
RA Villalón D.K., Muzny K.C., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahney J., Helton E., Kettman M., Madan A.C., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalish D.E.,
RA Schnerch A., Schein J.B., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC -!- FUNCTION: MAY PLAY A ROLE IN TUMOR PROGRESSION, METASTASIS
CC FORMATION AND TUMOR INVASION.

P48038 oryctolagus
Q9bqr3 homo sapien
Q9y5y6 homo sapien
P35036 anopheles g
P07338 rattus norv
Q9es87 rattus norv
P20918 mus musculu
P35041 anopheles g
P06867 sus scrofa
P03951 homo sapien
Q9esd1 mus musculu
P80009 canis famli

or send an email to license@isb-sib.ch.

```

CC CC EMBL; AJ429216; CAD22137.1; -
CC DR EMBL; AJ300738; CAC83350.1; -
CC DR HSSP; P00761; IANI.
CC DR MGD; MGI:2155445; Tmprss3.
CC DR InterPro; IPR001314; Chymotrypsin.
CC DR InterPro; IPR002172; LDL receptor A.
CC DR InterPro; IPR001254; Ser_protease_Try.
CC DR InterPro; IPR001190; Srdc_receptor.
CC DR Pfam; PF00057; ldl_recept_a; 1.
CC DR Pfam; PF00089; trypsin; 1.
CC DR PRINTS; PR00722; CHYMOTRYPSIN.
CC DR SMART; SM00192; LDLA; 1.
CC DR SMART; SM00202; SR; 1.
CC DR SMART; SM00202; Tryp_Spc; 1.
CC DR PROSITE; PS01209; LDLRA_1; 1.
CC DR PROSITE; PS00668; LDLRA_2; 1.
CC DR PROSITE; PS0287; SRCR_2; 1.
CC DR PROSITE; PS0240; TRYPSIN_DOM; 1.
CC DR PROSITE; PS00134; TRYPSIN_HIS; 1.
CC DR PROSITE; PS00135; TRYPSIN_SER; 1.
CC DR Hydroxase; Serine protease; Transmembrane; Signal-anchor; Zymogen;
CC DR Endoplasmic reticulum.
CC DR TRANSMEM 49 69
CC FT DOMAIN 70 453
CC FT DOMAIN 72 108
CC FT DOMAIN 104 205
CC FT DOMAIN 217 448
CC FT ACT_SITE 257 257
CC FT ACT_SITE 304 304
CC FT ACT_SITE 400 400
CC FT SITE 216 217
CC FT DISULFID 73 85
CC FT DISULFID 79 98
CC FT DISULFID 92 107
CC FT DISULFID 129 194
CC FT DISULFID 142 204
CC FT DISULFID 207 324
CC FT DISULFID 242 258
CC FT DISULFID 338 405
CC FT DISULFID 369 385
CC FT DISULFID 396 424
CC FT CARBOHYD 221 221
CC FT CONFLICT 117 117
CC FT CONFLICT 246 246
CC FT SEQUENCE 453 AA; 49491 MW; 1ABCBF10AF6E1BF6 CRC64;
Query Match 29.3%; Score 687; DB 1; Length 453;
Best Local Similarity 36.9%; Pred. No. 5e-52;
Matches 161; Conservative 72; Mismatches 153; Indels 50; Gaps 14;
CC QY 22 IPMETFRKVGIPPIIIALLSLASIIIVVILKVID---KYIFLCGQPLHFIPRKQLCDGE 78
CC DB 44 LPLKPPPIIIVIGIALLALA-----IGLIHFDCSGKY--RCHSFKCIELTRCDGV 95
CC QY 79 LDCPLGEDEHCVKSPPEGPAVARLSKDRSTLOVLDSTGNWFSACFNFTALAEATAC 138
CC DB 96 SDCKNADEYRC-----VRVSGRAALQVFTAAA--WRTWCSDDWKSHYAKIAC 142
CC QY 139 RMGYSS-----KPTFRVIEGPDQDLDVVEITENSQELMRNNSGFLCSGL 186
CC DB 143 AOLGFFSYSSDHLRVDALEEQQGDFSVINHLSDDKVTLAHHVYMRG---CTSGHV 199
CC QY 187 VSLHCLACG-KSLKTPRVVGGEEASVDSNPWQVSIQYDKHVCGGSIILDPHWLTAHCF 245
CC DB 200 VTLKCSACGTRGYSRIVGNNMSSITQMPWQSLQFGVHLCCGSGVITPLMIVTAHCV 259
CC QY 246 RKHTDVENKVRAGSKL--GSFPLSLAKIILIEFNPMY-PK--DNIDIALMKQLPPLTF 300
CC DB 260 YDLIYHKPSWTVOVGLVSLMDSVPVSHLVKKI---YHSKYKPKRLGNDIALMKLSEPLTF 316

```

```

QY 301 SCTVRPICLPFDEELTPATPLWILICGFTKQNGCKMSDILLOASVQVIDSTCNADDAY 360
DB 317 DETIQICLPINSEENFPDGKLCWTSGWGAT-EDGDGASPVLANHAAPLTSNKICNHRDVI 375
QY 361 QGEVTEKMCAGIPEGGVDTCCGDSGGLPMYQSDQ-WHVVGIVSWGCGGPRSTPGVTK 419
DB 376 GGIISPSMLCAGYLKGVDSQQDSGGPLVCERRLWKLVGATSGFICGAENVKPGVYTR 435
QY 420 VSAYLNNIYNVWKAEL 435
DB 436 ITSFLDWIHEQLERDL 451
RESULT 3
TMS2_HUMAN
ID TMS2_HUMAN STANDARD; PRT; 492 AA.
AC O1533; Q9EXX1;
DT 15-JUL-1998 (Rel. 36, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Transmembrane protease, serine 2 precursor (EC 3.4.21.-).
GN TMRSS2 OR PRSS10.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97468144; PubMed=9325052;
RA Paoloni-Giacobino A., Chen H., Peitsch M.C., Rossier C.,
RA Antonarakis S.E.;
RT "Cloning of the TMRSS2 gene, which encodes a novel serine protease
RT with transmembrane, LDLRA, and SRCR domains and maps to 21q22.3";
RL Genomics 44:309-320(1997).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=21309069; PubMed=11414763;
RA Teng D.H., Chen Y., Lian L., Ha P.C., Tavtigian S.V., Wong A.K.;
RT "Mutation analyses of 268 candidate genes in human tumor cell lines.";
RL Genomics 74:352-364(2001).
RN [3]
RP SEQUENCE FROM N.A., AND MUTAGENESIS.
RX MEDLINE=21139112; PubMed=11245484;
RA Afar D.E.H., Vivanco I., Hubert R.S., Kuo J., Chen E., Safran D.C.,
RA Raitano A.B., Jakobovits A.;
RT "Catalytic cleavage of the androgen-regulated TMRSS2 protease results
RT in its secretion by prostate and prostate cancer epithelia.";
RL Cancer Res. 61:1686-1692(2001).
RN [4]
RP TISSUE SPECIFICITY.
RX MEDLINE=21104370; PubMed=11169526;
RA Vaarala M.H., Porvari K.S., Kellokumpu S., Kyllonen A.P., Viikko P.T.;
RT "Expression of transmembrane serine protease TMRSS2 in mouse and
RT human tissues";
RL J. Pathol. 193:134-140(2001).
CC -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. ACTIVATED BY
CC CLEAVAGE AND SECRETED.
CC -!- TISSUE SPECIFICITY: EXPRESSED STRONGLY IN SMALL INTESTINE. ALSO
CC EXPRESSED IN PROSTATE, COLON, STOMACH, AND SALIVARY GLAND.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
CC -!- SIMILARITY: Contains 1 SRCR domain.
CC -!- SIMILARITY: Contains 1 LDL-receptor class A domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U75329; AAC51784.1; -

```

DR EMEL; AF123453; AAD37117.1; --
 DR EMEL; AF270487; AAK29280.1; --
 DR HSSP; P00763; IDPO.
 DR MEROPS; S01.247; --
 DR MIM; G02060; --
 DR GO; G0005887; C: integral to plasma membrane; TAG.
 DR InterPro; IPR001314; Chymotrypsin.
 DR InterPro; IPR002172; LDL_receptor A.
 DR InterPro; IPR001254; Ser_protease Try.
 DR Pfam; PF00089; trypsin; 1_receptor.
 DR PRINTS; PR00722; CHYMOTRYPSIN.
 DR SMART; SM00192; LDLA; 1.
 DR SMART; SM00202; SR; 1.
 DR SMART; SM00202; Tryp_Spc; 1.
 DR PROSITE; PS01209; LDLRA 1; 1.
 DR PROSITE; PS00668; LDLRA 2; 1.
 DR PROSITE; PS00420; SRCR 1; FALSE_NEG.
 DR PROSITE; PS00287; SRCR 2; 1.
 DR PROSITE; PS00240; TRYPSIN_DOM; 1.
 DR PROSITE; PS00134; TRYPSIN_HIS; 1.
 DR PROSITE; PS00135; TRYPSIN_SER; 1.
 DR Hydrolase; Serine protease; Signal-anchor; Zymogen;
 KW Polymorphism.
 KW CHAIN 1 255 TRANSMEMBRANE PROTEASE, SERINE 2, NON-
 FT CATALYTIC CHAIN.
 FT CHAIN 256 492 TRANSMEMBRANE PROTEASE, SERINE 2,
 FT CATALYTIC CHAIN.
 FT DOMAIN 1 84 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 85 105 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
 FT (POTENTIAL).
 FT DOMAIN 106 492 EXTRACELLULAR (POTENTIAL).
 FT DOMAIN 112 149 LDL-RECEPTOR CLASS A.
 FT DOMAIN 150 242 SRCR.
 FT DOMAIN 256 492 SERINE PROTEASE.
 FT ACT_SITE 296 296 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 345 345 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 441 441 CLEAVAGE (POTENTIAL).
 FT SITE 255 256 BY SIMILARITY.
 FT DISULFID 113 126 BY SIMILARITY.
 FT DISULFID 120 139 BY SIMILARITY.
 FT DISULFID 133 148 BY SIMILARITY.
 FT DISULFID 172 231 BY SIMILARITY.
 FT DISULFID 185 241 BY SIMILARITY.
 FT DISULFID 244 365 INTERCHAIN (BY SIMILARITY).
 FT DISULFID 281 297 BY SIMILARITY.
 FT DISULFID 410 426 BY SIMILARITY.
 FT DISULFID 437 465 BY SIMILARITY.
 FT CARBOHYD 213 213 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 249 249 N-LINKED (GLCNAC...) (POTENTIAL).
 FT VARIANT 449 449 K -> N (IN dbSNP:1056602).
 FT MUTAGEN 255 255 R->Q: LOSS OF CLEAVAGE.
 FT MUTAGEN 441 441 S->A: LOSS OF ACTIVITY.
 FT CONFLICT 160 160 M -> V (IN REF. 3).
 FT CONFLICT 242 242 I -> L (IN REF. 1).
 FT CONFLICT 329 329 E -> Q (IN REF. 1).
 FT CONFLICT 489 491 RAD -> KAN (IN REF. 1).
 FT SEQUENCE 492 AA; 53891 MW; CAB44FD174A9076B CRC64;
 Query Match 28.8%; Score 673.5; DB 1; Length 492;
 Best Local Similarity 38.8%; Pred. No. 8.3e-51;
 Matches 149; Conservative 58; Mismatches 128; Indels 49; Gaps 13;
 75 CDGELDCPLGDEDEHCVKSPPEGPAVAVRLSKDRSTLQVLDNATGNFACDFNFTEALA 134
 133 CDGVSHCPGDEENCRVLY-GP-----NFIQWISSQKSWHPVQDDWNNYNG 181
 135 ETACROMGYSSKPTTFAVIGPDQDLVVEITENSQELRMNMSG-----PCL 182
 182 RAACRDMGY--KXNFYSSQ-----GIVDDSGSTFMKLNISAGNVDIYKKLIYHSDACS 232

QY 183 SGLSVLSHCLACGKSL---KTPRVVGGEASVDSPWQVSIQYDKQHVCGSILDPHWL 239
 DB 233 SKAVVSLRCIACGVNLSSRQSVIGESALPANPWQVSLHVQNVHVCQGSITPEWIV 292
 QY 240 TAAHCFRKH-TDVFNKVRAGSKLGSF---PSLAVAKIIIIIEPNMY---PKDNIDIAL 291
 DB 293 TAAHCVCKPLNNPWHWTAFAGILR-QSFMFYAGYQVEKVI---SHPNYDSKTKNIDIAL 348
 QY 292 MKLOPPLTFSGTVRPLCLPFDEELTPATPLNIIHGFTKQNGKMSDILLOASVOVIDS 351
 DB 349 MKLQKPLTFNDLVKPLCPENPGMMLQPEQLCWISGWGATEEK-GKTSEVLNAAKVLIIET 407
 QY 352 TRCNADDAYQGEVTEKXMCAGIEGGVGDTCQDSGGFLMYQSDQ-WHVVGIVSWGYGCGG 410
 DB 408 QRCNSRYVYDNLITPAMICAGFLQGNVDSQDSGGFLVTSKKNIIWLLIGDTSWGSQCAK 467
 QY 411 PSTPGVYTKVSAYLWNIYNNWKA 434
 DB 468 AVREPGVYGNVMVFTDWIYEQMRAD 491
 RESULT 4
 TM83 HUMAN STANDARD; PRT; 454 AA.
 ID TM83 HUMAN AC P57727;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Transmembrane protease, serine 3 (EC 3.4.21.-) (Serine protease
 DE TAGD-12) (tumor associated differentially-expressed gene-12 protein).
 GN TMPSR3 OR TAGD12 OR ECHOS1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID:9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORMS A AND T).
 RC TISSUE=Ovarian carcinoma;
 RX MEDLINE=20521358; PubMed=11068177;
 RA Underwood L.J., Shigemasa K., Tanimoto H., Beard J.B., Schneider E.N.,
 RA Wang Y., Parmley T.H., O'Brien T.J.;
 RT "Ovarian tumor cells express a novel multi-domain cell surface serine
 RT protease";
 RL Biochim. Biophys. Acta 1502:337-350(2000).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORMS A, B AND D), AND VARIANT ILE-53.
 RX MEDLINE=20578749; PubMed=11137999;
 RA Scott H.S., Kudoh J., Wattenhofer M., Shibuya K., Berry A., Chraast R.,
 RA Guipponi M., Wang J., Kawasaki K., Asakawa S., Minoshima S.,
 RA Younus F., Mehdi S.Q., Radhakrishna U., Papasavvas M.P., Gehrig C.,
 RA Rossier C., Korostishevsky M., Gal A., Shimizu N., Bonne-Tamir B.,
 RA Antonarakis S.E.;
 RT "Insertion of beta-satellite repeats identifies a transmembrane
 RT protease causing both congenital and childhood onset autosomal
 RT recessive deafness";
 RL Nat. Genet. 27:59-63(2001).
 RN [3]
 RP SUBCELLULAR LOCATION, AND FUNCTION IN ENAC CLEAVAGE.
 RX PubMed=12393794;
 RA Guipponi M., Vuagniaux G., Wattenhofer M., Shibuya K., Vazquez M.,
 RA Dougherty L., Scamuffa N., Guida E., Okui M., Rossier C., Hancock M.,
 RA Buchet K., Raymond A., Hummler E., Marzella P.L., Kudoh J.,
 RA Shimizu N., Scott H.S., Antonarakis S.E., Rossier B.C.;
 RT "The transmembrane serine protease (TMPSR3) mutated in deafness
 RT DFNB8/10 activates the epithelial sodium channel (ENaC) in vitro";
 RL Hum. Mol. Genet. 11:2829-2836(2002).
 RN [4]
 RP VARIANTS DPB8/DFNB10 CYS-251 AND LEU-404.
 RX MEDLINE=21354482; PubMed=11462234;
 RA Masmoudi S., Antonarakis S.E., Schwede T., Ghorbel A.M., Gratri M.,
 RA Pappasavvas M.P., Drira M., Elgaied-Boullila A., Wattenhofer M.,
 RA Rossier C., Scott H.S., Ayadi H., Guipponi M.;
 RT "Novel missense mutations of TMPSR3 in two consanguineous Tunisian

InterPro: IPR001254; Ser protease Try.			
InterPro: IPR001190; Srcr receptor.			
Pfam: PF00057; ldl_recept_a.1.			
Pfam: PF00039; tryptsin_1.			
PRINTS; PR00722; CHYMOTRYPSIN.			
SMART; SM00192; LDLA; 1.			
SMART; SM00202; SR; 1.			
SMART; SM00202; TRYSPG; 1.			
PROSITE; PS01209; LDLRA_1; 1.			
PROSITE; PS00658; LDLRA_2; 1.			
PROSITE; PS00420; SRCR_1; FALSE_NEG.			
PROSITE; PS0287; SRCR_2; 1.			
PROSITE; PS0240; TRYPSIN_DOM; 1.			
PROSITE; PS00134; TRYPSIN_HIS; 1.			
PROSITE; PS00135; TRYPSIN_SER; 1.			
Hydrolase; Serine protease; Transmembrane; Signal-anchor; Zymogen;			
Endoplasmic reticulum; Deafness; Alternative splicing;			
Disease mutation; Polymorphism.			
DOMAIN 1 48			
CYTOPLASMIC (POTENTIAL).			
SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)			
(POTENTIAL).			
EXTRACELLULAR (POTENTIAL).			
LDL-RECEPTOR CLASS A.			
SRCR.			
SERINE PROTEASE.			
CHARGE RELAY SYSTEM (BY SIMILARITY).			
CHARGE RELAY SYSTEM (BY SIMILARITY).			
CHARGE RELAY SYSTEM (BY SIMILARITY).			
CLEAVAGE (POTENTIAL).			
BY SIMILARITY.			
BY SIMILARITY.			
BY SIMILARITY.			
BY SIMILARITY.			
BY SIMILARITY.			
BY SIMILARITY.			
BY SIMILARITY.			
BY SIMILARITY.			
N-LINKED (GLCNAC...) (POTENTIAL).			
Missing (in isoform B).			
/Fttd=vsp_005391.			
EMIPVCLPSEENFPDGKVGTSWGATEDGADGASPVLIN			
HAAVPLSNKTCNHRDVGITISPSMLCAGLGGVDSQQG			
DSGGLPVCOERBLKLVGATSGFGIGCAEVNKPVGVTYVTSF			
LDWHEQMERDLKT -> GTSGSLCGSAALPLQEDLQLLI			
EAPL (in isoform D).			
/Fttd=vsp_005392.			
DLVLPKSTIQGVLSLDLNPAPSHLVKEVIYH -> EIVA			
PREARADRGRLKLCWRKPTKMGKPRPSHS (in isoform			
T).			
/Fttd=vsp_005393.			
Missing (in isoform T).			
/Fttd=vsp_005394.			
V -> I.			
/Fttd=var_010781.			
D -> G (in DFNB8/DFNE10).			
/Fttd=var_013490.			
R -> W (in DFNB8/DFNE10).			
/Fttd=var_013491.			
G -> S.			
/Fttd=var_013492.			
D -> N.			
/Fttd=var_013493.			
C -> F (in DFNB8/DFNE10).			
/Fttd=var_013494.			
W -> C (in DFNB8/DFNE10).			
/Fttd=var_011678.			
I -> V (in dBSNP:2839500).			
/Fttd=var_013101.			
P -> L (in DFNB8/DFNE10).			
/Fttd=var_011679.			
C -> R (in DFNB8/DFNE10).			
DOMAIN 49 69			
TRANSMEM			
DOMAIN 70 454			
DOMAIN 72 108			
DOMAIN 109 205			
DOMAIN 217 454			
ACT_SITE 257 257			
ACT_SITE 304 304			
ACT_SITE 401 401			
SITE 216 217			
DISULFID 73 85			
DISULFID 79 98			
DISULFID 92 107			
DISULFID 129 194			
DISULFID 142 204			
DISULFID 207 324			
DISULFID 242 258			
DISULFID 338 407			
DISULFID 370 386			
DISULFID 397 425			
CARBOHYD 221 221			
VARSPPLIC 1 127			
VARSPPLIC 318 454			
VARSPPLIC 261 293			
VARSPPLIC 294 454			
VARIANT 53 53			
VARIANT 103 103			
VARIANT 109 109			
VARIANT 111 111			
VARIANT 173 173			
VARIANT 194 194			
VARIANT 251 251			


```

46  -----SDQB-----PYPQVSSADARLMVFDKTECTWELLCSSRNARVAGLSC 90
139 ROMGYSSKPTFRAVEIGPDODLDVVEITEN-----SOELMRNSSGPGCLS 183
91  EEMGF-----LRAL---THSELDVRTAGANGTSFGFCVDEGRLPHTQRLLEVLISVCDPR 142
184 GSVLSHLCLAG-KSLKTPRVVGEASVDSWPQVQSIQVDKHQVCGGSLDPHWLVTAA 242
143 GRFLAALICODGRRKLPVDRIVGRDTSLGKRPWPQVSLRYDGAHLGGLSLGDPWLTA 202
243 HCRKHTDVEN-WKVRAGSKLGSFSPSLAVAKIILIEFNMPY-----KNDIALMK 293
203 HCFPERNVLRSRVRPAGVAGAPHLQGLQGVAVYHGGVLPFRDPNSSENSNDIALVH 262
294 LQPLTPSGTVRPICLPFPPDELPATPLMIIGWFTKQNGKMSDILLOASVOVIDSTR 353
263 LSEPLPTEVIQVCLPAAGALVDGKICTVTGNGT-QYVGGQAGVLQEARVPIISNDV 321
354 CNADDAYGVEVTERQMCAGIPEGWDTCCQDSGGPLMYQ-----SDQHVVVGIIVSWGYC 408
322 CNGADFQNGIKPMFCAGYPEGIDACQDSGGFFVCEDSISRTPRWRLCGIVSGWTGC 381
409 GGPSTGTVTKVSAYLNNIYNWK 432
382 ALAQKPGVYTKVSDFREWIPOAIK 405

RESULT 8
NTK HUMAN
D -ENTK_HUMAN STANDARD; PRT; 1019 AA.
C P98073;
T 01-FEB-1996 (Rel. 33, Created)
T 01-FEB-1996 (Rel. 33, Last sequence update)
T 15-SEP-2003 (Rel. 42, Last annotation update)
E Enteropeptidase precursor (EC 3.4.21.9) (Enterokinase).
N PRS7 OR ENTK.
S Homo sapiens (Human).
C Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
C Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
X NCBI_TaxID=9606;
P [1]
SEQUENCE FROM N.A.
C TISSUE=DuoDenum;
C MEDLINE=95234679; PubMed=7718557;
A Kitamoto Y., Veille R.A., Donis-Keller H., Sadler J.E.;
T "cDNA sequence and chromosomal localization of human enterokinase,
T the proteolytic activator of trypsinogen.";
T Biochemistry 34:4562-4568 (1995).
[2]
SEQUENCE FROM N.A., AND DISEASE.
MEDLINE=21606074; PubMed=11719902;
A Holzinger A., Maier E.M., Buck C., Mayerhofer P.U., Kappler M.,
A Haworth J.C., Moroz S.P., Buck H.-B., Sadler J.E., Roscher A.A.;
T "Mutations in the proenteropeptidase gene are the molecular cause of
T congenital enteropeptidase deficiency.";
T Am. J. Hum. Genet. 70:20-25 (2002).
[3]
SEQUENCE FROM N.A.
MEDLINE=20289799; PubMed=10830953;
A Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,
A Park H.-S., Toyoda A., Ishii K., Totoki Y., Choi D.-K., Groner Y.,
A Soeda E., Ohki M., Takagi T., Sakaki Y., Taudien S., Blechschmidt K.,
A Polley A., Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D.,
A Reichwald K., Rump A., Schillhabel M., Schudy A., Zimmermann W.,
A Reenthald A., Kudoh J., Shibuya K., Kawasaki K., Asakawa S.,
A Shintani A., Sasaki T., Negamine K., Mitsuyma S., Antonarakis S.E.,
A Miroshima S., Shimizu N., Nordisiek G., Hornischer K., Brandt P.,
A Schaefer M., Schoen O., Desario A., Reichelt J., Kauer G., Bloecker H.,
A Ranser J., Beck A., Klages S., Hennig S., Riesselmann L., Dagand E.,
A Wehrmeyer S., Borzym K., Gardiner K., Nizetic D., Francis F.,
A Lehrach H., Reinhardt R., Yaspo M.-L.;
T "The DNA sequence of human chromosome 21."

```


FT	ACT_SITE	222	222		CHARGE RELAY SYSTEM (BY SIMILARITY).
FT	ACT_SITE	276	276		CHARGE RELAY SYSTEM (BY SIMILARITY).
FT	ACT_SITE	372	372		CHARGE RELAY SYSTEM (BY SIMILARITY).
FT	DIGULFID	172	296		INTERCHAIN (BY SIMILARITY).
FT	DIGULFID	207	223		BY SIMILARITY.
FT	DIGULFID	341	357		BY SIMILARITY.
FT	DIGULFID	368	400		BY SIMILARITY.
FT	CARBONYD	131	131		N-LINKED (GLCNAC..) (POTENTIAL).
FT	VARSPLIC	25	44		/Ftid=VSP_007232.
FT	CONFLICT	85	85		L -> F (IN REF. 2 AND 3).
FT	CONFLICT	204	204		T -> Y (IN REF. 3).
FT	CONFLICT	214	214		G -> R (IN REF. 3).
FT	CONFLICT	228	229		NR -> ET (IN REF. 3).
FT	CONFLICT	264	264		P -> L (IN REF. 3).
FT	CONFLICT	281	281		H -> N (IN REF. 3).
SQ	SEQUENCE	436 AA;	46787 MW;	4A0993148C620BD0 CRC64;	
 Query Match 24.7%; Score 578; DB 1; Length 436;					
Best Local Similarity 31.1%; Pred.No.1.4e-42;					
Matches 136; Conservative 62; Mismatches 150; Indels 90; Gaps 10					
Qy	30	VGIPIIIALLSASIIIVVLIKYLVKCYFLGCPQLHFIPRKQLCDGELDCPLGEDEH	89	:::::	
Db	42	VGTLFUTGTGAASWAVTILLQ	68	-----SDQE-	
Qy	90	CVKSFPFGPAVARLSKDRSTLOVLDSATGNWFSACDFNFTEALAEATACROMGY	143	-----	
Db	69	-----PLYQVQLSPGDSRLAVLDKTEGWRLLCSSRSNARVAGLGCCEWGFLRALAH	120	:::::	
Qy	144	-----SSKPTFRVAIIG	189	-----PDQLDVVEITENSQELMRNSSGGCLSGSLVSL	
Db	121	SELDFRTAGANGTSGFPCVDEGGGLPLAQRLDDVISVCD	167	-----CPRGRFLTA	
Qy	190	HCLAAG-KSLKTRPVVGGBEASVDSPWQYSIQYDKOHVCGSILDPHVWLTAACHCKH	248	:::::	
Db	168	TCQDCGERKLPIVRIVCGQSISGRWPQWSLRVDGTHLCGSLSDGDVNLTAACHCFER	227	-----	
Qy	249	TDVFN-WKVVRAGSKLGSPFSFLAVAKIIITEFNMPY	299	-----KNDIALMKLQPFLLT	
Db	228	NRVLSRWRFVAGAVARTSPHAVLGVOAVIHGGYLPRDPDTIDNSNDIALVHLSSLP	287	-----	
Qy	300	PSGTVRPICLPFDBELTPATPIWIIGWTTKQNGKGKSIDLLQASVQVIDSTRCNADDA	359	-----	
Db	288	LTYEIOPVCLFAAGQALVDGKVTCTTGWNT-QPYGOAMVLQEARVPILSIIEVCNSPDF	346	-----	
Qy	360	YOGEVTEKMWKACGIPGGGVDTCCGDSGGELMYQ	414	-----SDWHVVGVHSXGYGCGGSGTP	
Db	347	YGNQIKPKFCAGYPPEGIDACGDSGGPFVCSDSISGTSWRWLCGIVSWGTCALARKP	406	-----	
Qy	415	GVYTKVSAYLNWTINYWK	432	-----	
Db	407	GVYTKVTFREWIFKAIK	424	-----	
 RESULT 10					
ID	TMS5 HUMAN	STANDARD;	PRTI	457 AA.	
AC	Q9H3S3;				
DT	16-OCT-2001 (Rel. 40, Created)				
DT	16-OCT-2001 (Rel. 40, Last sequence update)				
DT	28-FEB-2003 (Rel. 41, Last annotation update)				
DE	Transmembrane protease, serine 5 (EC 3.4.21.-) (Spinesin).				
GN	TMPRS55.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OX	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
NCBI_Taxid=9606;	[1]				
RN	SEQUENCE FROM N.A.				
RP	TISSUE=Brain;				
RC	PubMed=11741986;				
RA	Yamaguchi N., Okui A., Yamada T., Nakazato H., Mitsui S.;				

HEPSIN RAT STANDARD; PRT; 416 AA.

Q05511; 01-FEB-1994 (Rel. 28, Created)

01-FEB-1994 (Rel. 28, Last sequence update)

15-SEP-2003 (Rel. 42, Last annotation update)

Serine protease hepsin (BC 3.4.21.-).

HPN.

Rattus norvegicus (Rat).

Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

NCBI_TaxID=10116;

[1]

SEQUENCE FROM N.A.

TISSUE=Liver;

MEDLINE=93305733; PubMed=8318546;

Farley D., Raymond F., Nick H.;

"Cloning and sequence analysis of rat hepsin, a cell surface serine proteinase."

Biochim. Biophys. Acta 1173:350-352(1993).

-!- FUNCTION: Plays an essential role in cell growth and maintenance of cell morphology.

-!- SUBCELLULAR LOCATION: Type II membrane protein.

-!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; X70900; CAA50256.1; --

PIR; S33777; S33777.

HSSP; P00763; IDPO.

MEROPS; S01.224; --

InterPro; IPR001314; Chymotrypsin.

InterPro; IPR001254; Sex protease Try.

InterPro; IPR001190; Src1_receptor.

Pfam; PF00089; trypsin; 1.

PRINTS; P000722; CHYMOTRYPSIN.

SMART; SM00202; SR; 1.

SMART; SM00020; Tryp_SPC; 1.

PROSITE; PS00240; TRYP SIN DOM; 1.

PROSITE; PS00134; TRYP SIN_HIS; 1.

PROSITE; PS00335; TRYP SIN_SSR; 1.

Hydrolase; Serine protease; Transmembrane; Signal-anchor.

CHAIN 1 161

SRINE PROTEASE HEP SIN, NON-CATALYTIC CHAIN (POTENTIAL).

CHAIN 162 416

SRINE PROTEASE HEP SIN, CATALYTIC CHAIN (POTENTIAL).

DOMAIN 1 16

TRANSMEM 17 43

CYTOPLASMIC (POTENTIAL).

SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN) (POTENTIAL).

EXTRACELLULAR (POTENTIAL).

SERINE PROTEASE

CHARGE RELAY SYSTEM (BY SIMILARITY).

CHARGE RELAY SYSTEM (BY SIMILARITY).

CHARGE RELAY SYSTEM (BY SIMILARITY).

INTERCHAIN (BY SIMILARITY).

BY SIMILARITY.

BY SIMILARITY.

BY SIMILARITY.

N-LINKED (GLCNAC...) (POTENTIAL).

SEQUENCE 416 AA; 44926 MW; E5A9F8FA950E180 CRC64;

Query Match 24.3%; Score 568.5; DB 1; Length 416;

Best Local Similarity 30.5%; Pred. No. 8.e-42;

Matches 137; Conservative 64; Mismatches 155; Indels 93; Gaps 11;

19 KPRIPWETKRGVIFIIALLSLASIIIVVLLKYLKDYFLCQPLHFIPKQLCDGE 78

14 RPKVAALT---VGLTLFLTGIGAASWALITLLR----- 44

QY 79 LDCPLGEDDEHCVKSPFEGPVAVALRSKDRSLQVLDSATGNWFSACFDNFTALAEATC 138

DB 45 -----SDQE-----FLYQVQLSPGDSLLVLDKTEGTWRLLCSSRSNARVAGLCG 89

QY 139 ROMGY-----SSKPTFRVETG-----PDQLDVEITENSDELKRNSS 178

DB 90 BEMGFIRALAHSELVDRTAGANGTSGFFCVDRGGLFLAQRLDLDVISVC----- 138

QY 179 GFLCSGLSVLHCLACG-KSLKTPRVVGEEASVDSWPQVSIQYDKQHVCGSGILDPHW 237

DB 139 ---CPRGRLTATCQGRKLPVDRIVGQDSGLGRWPQVSLRYDTHLCGSGLLSGDW 196

QY 238 VLTAACFRKHTDVEN-MKVRAGSDKLSGFFSLAVAKIIIEFNMPY-----KND 288

DB 197 VLTAACFPERNRLSRNRPAGAVARTSPHVAQLGVQAVIYHGGVLPDRPTIDENSND 256

QY 289 IALMKLQPLTFSGTVTRPCLPFPFDEELTPATPLMIWGFTKQNGKMSDILLOASVOV 348

DB 257 IALVHLSLPLTEYIQVCLPFAQQAALVDGKVTGTGNGT-QFYQQAVALQEARVPI 315

QY 349 IDSTRCNADDAYQGEVTEKMCAGIPEGVDTCQDGGGELMYQ-----SDQHVVGIVS 403

DB 316 ISNVCNSPDPFYGNQIKPKMFCAGYPEGIDACQDGGHFCVEDRISGTSRWRLCGIVS 375

QY 404 WYGGGSPSTPGVYTKVAYLNIWVNVK 432

DB 376 WGTGALARKPGVYTKVIDFREWIFQAIK 404

RESULT 13

HATT HUMAN

ID HATT HUMAN STANDARD; PRT; 418 AA.

AC O60235;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Airway trypsin-like protease precursor (BC 3.4.21.-).

GN HAT.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

SEQUENCE FROM N.A.

RX MEDLINE=98234382; PubMed=9565616;

RA Yamaoka K., Masuda K.-I., Ogawa H., Takagi K., Umemoto N., Yamaoka S.;

RT "Cloning and characterization of the cDNA for human airway trypsin-like protease."

RT J. Biol. Chem. 273:11895-11901(1998).

RL [2]

SEQUENCE OF 187-206, AND CHARACTERIZATION.

RX MEDLINE=97224034; PubMed=9070615;

RA Yamaoka S., Ohnishi T., Kawano S., Tsuchihashi S., Ogawara M., Masuda K.-I., Yamaoka K., Takahashi M., Sano T.;

RT "Purification, characterization, and localization of a novel trypsin-like protease found in the human airway."

RL Am. J. Respir. Cell Mol. Biol. 16:300-308(1997).

CC -!- FUNCTION: May play some biological role in the host defense system on the mucous membrane independently of or in cooperation with other substances in airway mucous or bronchial secretions.

CC -!- CATALYTIC ACTIVITY: Preferentially cleaves the C-terminal side of arginine residues at the P1 position of certain peptides, cleaving Boc-Phe-Arg-4-methylcoumaryl-7-amide most efficiently and having an optimum pH of 8.6 with this substrate.

CC -!- ENZYME REGULATION: Strongly inhibited by diisopropyl fluorophosphate, leupeptin, antipain, aproinin, and soybean trypsin inhibitor, but hardly inhibited by secretory leukocyte protease inhibitor at 10 microM.

CC -!- SUBUNIT: Monomer.

CC -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. ACTIVATED BY CLEAVAGE AND SECRETED.

CC -!- TISSUE SPECIFICITY: Located in the cells of the submucosal serous


```
DR DR InterPro; IPR002172; LDL_receptor_A.
DR DR InterPro; IPR000998; MAM domain.
DR DR InterPro; IPR00082; SEA domain.
DR DR InterPro; IPR001254; Ser_protease_Try.
DR DR InterPro; IPR001190; Srcr_receptor.
DR DR Pfam; PF00431; CUB; 2.
DR DR Pfam; PF00057; ldl_recept_a; 2.
DR DR Pfam; PF00629; MAM; 1.
DR DR Pfam; PF01390; SEA; 1.
DR DR Pfam; PF00530; SRCR; 1.
DR DR Pfam; PF00089; trypsin; 1.
DR DR PRINTS; PR00722; CHYMOTRYPSIN.
DR DR PRINTS; PR00261; LDLRECEPTOR.
DR DR PRINTS; PR00020; MAMDOMAIN.
DR DR SMART; SM00042; CUB; 2.
DR DR SMART; SM00192; LDLA; 2.
DR DR SMART; SM00137; MAM; 1.
DR DR SMART; SM00200; SEA; 1.
DR DR SMART; SM00202; SR; 1.
DR DR SMART; SM00020; Tryp_Spc; 1.
DR DR PROSITE; PS01180; CUB; 2.
DR DR PROSITE; PS01209; LDLRA_1; 2.
DR DR PROSITE; PS00068; LDLRA_2; 2.
DR DR PROSITE; PS00740; MAM_1; 1.
DR DR PROSITE; PS00600; MAM_2; 1.
DR DR PROSITE; PS00024; SEA; 1.
DR DR PROSITE; PS00420; SRCR_1; FALSE_NEG.
DR DR PROSITE; PS00287; SRCR_2; 1.
DR DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR DR PROSITE; PS00135; TRYPSIN_SER; 1.
DR DR Signal-anchor; Glycoprotein; Hydrolase; Serine protease; Zymogen;
DR DR Transmembrane; Repeat.
DR CHAIN 1 829
DR CHAIN 830 1069
DR DOMAIN 1 18
DR TRANSMEM 19 47
DR DOMAIN 48 1069
DR DOMAIN 52 169
DR DOMAIN 227 268
DR DOMAIN 270 379
DR DOMAIN 387 549
DR DOMAIN 569 679
DR DOMAIN 686 724
DR DOMAIN 723 816
DR DOMAIN 830 1069
DR ACT_SITE 874 874
DR ACT_SITE 925 925
DR ACT_SITE 1021 1021
DR DISULFID 229 242
DR DISULFID 236 255
DR DISULFID 249 266
DR DISULFID 688 700
DR DISULFID 695 713
DR DISULFID 707 722
DR DISULFID 817 945
DR DISULFID 859 875
DR DISULFID 959 1027
DR DISULFID 991 1006
DR DISULFID 1017 1045
DR CARBOHYD 147 147
DR CARBOHYD 197 197
DR CARBOHYD 212 212
DR CARBOHYD 373 373
DR CARBOHYD 380 380
DR CARBOHYD 433 433
DR CARBOHYD 515 515
DR CARBOHYD 579 579
DR CARBOHYD 675 675
DR CARBOHYD 727 727
DR CARBOHYD 751 751
DR CARBOHYD 770 770
FT CARBOHYD 791 791 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 897 897 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 936 936 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 999 999 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1069 AA; 118735 MW; E62549E463743CD CRC64;
Query Match 23.6%; Score 552.5; DB 1; Length 1069;
Best Local Similarity 32.5%; Pred. No. 7.5e-40;
Matches 124; Conservative 71; Mismatches 150; Indels 37; Gaps 12;
QY 69 IPRKQLCDGELDCPLGDEBEHCVKSPFEGPAVAVRLSKDRSTLQVLD-SATGNWFSACFD 127
Db IPLGNLDCDSYPHCRDGSDEASCVRF-----LNGTRSNGLVQFNHISWHIACAE 750
QY 128 NPTALAEACROMGYSSKPTFRAVEI---GPDQDLVVEITENSQELRMENSSGPCLSG 184
Db NPTQISNEVCHLLGLGSANSMFISITGGP-----FVRVQAPNGSLIITPSLQCSQD 805
QY 185 SLVSLHC--LACGKSLKT---PRVVGEEASVSWPMQVSIQYDKQH---VCGGSILD 234
Db SLILLQCNHKSCEKKVTQKVPKIVGSDAQAGAWPWVVALYHRDRSTDLLOGASLVS 865
QY 235 PHWLTAAHC-FRKHTDVFHWKVRAGSKLGSFSLAVAKIII--IEFNKYP---KND 288
Db SDWLVSAAHCVYRRNLDFTRTAVLGHMQSNLTPQVYRVVDQIVINPHYDRRRKYND 925
QY 289 IALMKLOPLTFSGTVRPICLPFPDELTATPLWIIWGFTKQNGKMSDILLOASVQV 348
Db IAWHLEFKVNYTDYIQICLPENQIFIGRTCSIAGWYDKINAGSTVDVLEADVPL 985
QY 349 IDSTRCNADDAIQGEVTEKMKACIGPEGVDVTCGSDGGGLMYQ-SDQHWVGVISWGYG 407
Db ISNEKCO-QQLPEYNITESMICAGYERGGIDSCQDGGGLMCMQENNRFLVGVTSFGVQ 1044
QY 408 CGGPTGTVTKVSAYLNWYN 429
Db CALPNHPGVYRVVSQFIEWHS 1066
RESULT 15
ENTK_PIG STANDARD; PRT; 1034 AA.
AC P98074;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Enteropeptidase precursor (EC 3.4.21.9) (Enterokinase).
GN PRS7 OR ENTK
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RC SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=94327548; PubMed=8051081;
RA Matsushima M., Ichinose M., Yahagi N., Kakei N., Tsukada S.,
RA Miki K., Kurokawa K., Tashiro K., Shiohawa K., Shinomiya K.,
RA Uneyama H., Inoue H., Takahashi T., Takahashi K.;
RT "Structural characterization of porcine enteropeptidase.";
RL J. Biol. Chem. 269:19976-19982(1994).
CC -!- FUNCTION: RESPONSIBLE FOR INITIATING ACTIVATION OF PANCREATIC
CC PROTEOLYTIC PROENZYMES (TRYPSIN, CHYMOTRYPSIN AND CARBOXYPEPTIDASE
CC A). IT CATALYZES THE CONVERSION OF TRYPSINOGEN TO TRYPSIN WHICH IN
CC TURN ACTIVATES OTHER PROENZYMES INCLUDING CHYMOTRYPSINOGEN,
CC PROCARBOXYPEPTIDASES, AND PROELASTASES.
CC -!- CATALYTIC ACTIVITY: Selective cleavage of 6-Lys-|-Ile-7 bond in
CC trypsinogen.
CC -!- SUBUNIT: HETEROTRIMER OF A CATALYTIC (LIGHT) CHAIN, A MULTIDOMAIN
CC (HEAVY) CHAIN, AND A MINI CHAIN
CC -!- SUBCELLULAR LOCATION: Type II membrane protein (Probable).
CC -!- FM: THE CHAINS ARE DERIVED FROM A SINGLE PRECURSOR THAT IS
CC CLEAVED BY A TRYPSIN-LIKE PROTEASE.
```

Search completed: November 23, 2003, 07:49:48
Job time : 20.7538 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 23, 2003, 07:45:10 ; Search time 65.2201 Seconds
(without alignments)
1721.140 Million cell updates/sec

Write: US-09-607-745-2

Perfect score: 2342

Sequence: 1 MDPDSQDLNGLDKPLRKP.....VTKVSAYLNWYNNWKAEL 435

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 23.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phase.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_virus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	2167	92.5	405	4	Q6E86
2	1844	78.7	435	11	Q8VCA5
3	704.5	30.1	471	11	Q8CFE0
4	695.5	29.7	481	4	Q9BYE2
5	687	29.3	453	11	Q8K1T0
6	686.5	29.3	537	4	Q8YB1
7	686.5	29.3	767	13	Q9DGR2
8	681	29.1	453	11	Q8VDE0
9	670.5	28.6	492	4	Q96T73
10	646	27.6	490	11	Q920K3
11	636.5	27.2	445	11	Q8CJ17
12	631.5	27.0	455	11	Q8CDR0
13	612	26.1	371	11	Q8CJ16
14	550.5	23.5	799	11	Q9DB10
15	545	23.3	777	11	Q8CAN9
16	536	22.9	417	11	Q8VHJ4

ALIGNMENTS

RESULT 1

Q6E86	ID	Q96E86	PRELIMINARY;	PRT;	405 AA.
AC	Q96E86;				
DT	01-DEC-2001	(TREMBLrel. 19, Created)			
DT	01-DEC-2001	(TREMBLrel. 19, Last sequence update)			
DT	01-MAR-2003	(TREMBLrel. 23, Last annotation update)			
DE	Similar to transmembrane protease, serine 4 (Fragment).				
OS	Homo sapiens (Human)				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
OX	NCBI_TaxID=9606;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=Pancreas;				
RA	Strausberg R.;				
RL	Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.				
CC	-!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.				
DR	EMBL;	BC012752; AAH12752.1; -			
DR	HSP;	P00761; IAN1.			
DR	MEROPS;	S01.034; -			
DR	InterPro;	IPR001314; Chymotrypsin.			
DR	InterPro;	IPR002172; LDL receptor A.			
DR	InterPro;	IPR001254; Ser. protease Try.			
DR	Pfam;	PF00057; ldl recept_a; 1.			
DR	Pfam;	PF00089; tryptsin; 1.			
DR	PRINTS;	PRO0722; CHYMOTRYPSIN.			
DR	SMART;	SM00192; LDLa; 1.			
DR	SMART;	SM00202; SR; 1.			
DR	SMART;	SM00202; Tryp. spc; 1.			
DR	PROSITE;	PS0287; SSCR 2; 1.			
DR	PROSITE;	PS0240; TRYPSIN_DOM; 1.			
DR	PROSITE;	PS00134; TRYPSIN_HIS; 1.			
DR	PROSITE;	PS00135; TRYPSIN_SER; 1.			
KW	Hydrolase; Protease; Serine protease; Transmembrane.				
FT	NON TER 1				
SEQ	SEQUENCE 405 AA; 44474 MW; 951ACDE2D9D48E04 CRC64;				

Q8IUE2 homo sapien
Q8IU80 homo sapien
Q9CZ74 rattus norv
Q8VDV1 mus musculu
Q8VHK8 mus musculu
Q9GL10 oviss aries
Q9L674 xenopus lae
Q8R0P5 mus musculu
Q8BHM9 mus musculu
Q8N171 homo sapien
Q8BIK6 mus musculu
P79343 bos taurus
Q8BZ10 mus musculu
Q97506 sus scrofa
Q60491 cavia porce
Q8BM10 mus musculu
Q9R0W3 rattus norv
Q8BZ30 mus musculu
Q8BZ13 mus musculu
Q9PVX7 xenopus lae
Q8NF86 homo sapien
Q96RZ8 homo sapien
Q8DGR1 xenopus lae
Q29015 sus sp. pre
Q9V4N6 drosophila
Q9XZM7 strongyloce
Q9JJ17 rattus norv
Q9W7P9 paralichthy
Q9QYZ9 mus musculu

```
Query Match          92.5%; Score 2167; DB 4; Length 405;
Best Local Similarity 100.0%; Pred. No. 1.9e-202;
Matches 402; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

34 IIIALLSLASIIIVVLLIKVILDKYFFLCGQPLHPIPRKQICDGLDCEDEHCVKS 93
4 IIIALLSLASIIIVVLLIKVILDKYFFLCGQPLHPIPRKQICDGLDCEDEHCVKS 63

94 PPEGPAVAVRLSKDRSTLOVLDSATGNWFSACFDNFTALAEACROMGYSSKPTFRAVE 153
64 PPEGPAVAVRLSKDRSTLOVLDSATGNWFSACFDNFTALAEACROMGYSSKPTFRAVE 123

154 IGPDDLDWEITENSQELRMNNSGPPCLSGSLVSLHCLACSKSLKTPRVVGGEEASVDS 213
124 IGPDDLDWEITENSQELRMNNSGPPCLSGSLVSLHCLACSKSLKTPRVVGGEEASVDS 183

214 WPMQVSIQYDKQHVCGGSLDHPHWLTAACFRKHTDVFNNKVRAGSKLGSFPFLAVAK 273
184 WPMQVSIQYDKQHVCGGSLDHPHWLTAACFRKHTDVFNNKVRAGSKLGSFPFLAVAK 243

274 IIIIEFNPMYKDNIDIALMKLQFPPLTFSGTVRPICLPFFDEBELTPATPLWIGWFTKON 333
244 IIIIEFNPMYKDNIDIALMKLQFPPLTFSGTVRPICLPFFDEBELTPATPLWIGWFTKON 303

334 GKKSDIILQASVQVIDSTRCNADDAVQGEVTEKMCAGIPGGVDTCCGSGGGLMYTOS 393
304 GKKSDIILQASVQVIDSTRCNADDAVQGEVTEKMCAGIPGGVDTCCGSGGGLMYTOS 363

394 DQWVVGVISWGYCGGSPSTPGVYTKVSAYLNIWNVKAEI 435
364 DQWVVGVISWGYCGGSPSTPGVYTKVSAYLNIWNVKAEI 405

RESULT 2
38VCAS
ID Q8VCA5; PRELIMINARY; PRT; 435 AA.
AC Q8VCA5;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Similar to transmembrane protease, serine 4 (Channel-activating
DE protease 2).
EN TMRPS84.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
[1]
SEQUENCE FROM N.A.
TI Tissue=Breast tumor;
RA Strausberg R.;
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
[2]
SEQUENCE FROM N.A.
MEDLINE=22144321; PubMed=12149280;
Vugniaux G., Vallet V., Jaeger N.F., Hummler E., Rossier B.C.;
RT "Synergistic Activation of ENaC by Three Membrane-bound Channel-
RT activating Serine Proteases (mCAP1, mCAP2, and mCAP3) and Serum- and
RT Glucocorticoid-regulated Kinase (Sgk1) in Xenopus Oocytes.";
J. Gen. Physiol. 120:191-201(2002).
RL -I- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
EMBL; BC021368; AAH21368.1; -.
EMBL; AY043240; AAK95307.1; -.
HSSP; P00761; 1AN1.
MEROPS; S01.034; -.
MGD; MGI:2384877; Tmprs84.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR002172; LDL_receptor_A.
DR InterPro; IPR001254; Ser_protease_Try.
DR InterPro; IPR001190; Srcr_receptor.
DR Pfam; PF00057; ldl_recept_a; 1.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.

DR PROSITE; PS0287; SRCR 2; 1.
DR PROSITE; PS0240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Protease; Serine protease; Transmembrane.
SQ SEQUENCE 435 AA; 47495 MW; DC52E45A43E01369 CRC64;

Query Match          78.7%; Score 1844; DB 11; Length 435;
Best Local Similarity 77.2%; Pred. No. 6.4e-171;
Matches 336; Conservative 43; Mismatches 56; Indels 0; Gaps 0;

QY 1 MDPSDQPLNSLDVVKPRIPMETFRKVGIPPIIALLSLASIIIVVLLIKVILDKYFF 60
DB 1 MESDGGQPLNRRDIVPFRKPRRPOETFKVGIPPIIALLSLASIIIVVLLIKVILDKYFF 60

61 LCGQPLHPIPRKQICDGLDCEDEHCVKSPPEGPAVAVRLSKDRSTLOVLDSATGN 120
61 ICGSPLTFIQRGQCDGHLDCAEGDEBEHCYKDPFKPGVAVRLSKDRSTLOVLDATGT 120

121 WFSACFDNFTALAEACROMGYSSKPTFRAVEIGPDQDLDDVVEITENSQELRMNNSGP 180
121 WASVCFDNFTALAEACROMGYSDSQAFRAVEIRPQNLPVAQVTGNSQELQVNGSRS 180

181 CLSSSLVSLHCLACSKSLKTPRVVGGEEASVDSWPMQVSIQYDKQHVCGGSLDHPHWL 240
181 CLSSSLVSLRCLDCCKSLKTPRVVGGVEAPVDSWPMQVSIQYDKQHVCGGSLDHPHWL 240

241 AAHCFRKHDTDFNNKVRAGSKLGSFPFLAVAKIIIEFNPMYKDNIDIALMKLQFPPLTF 300
241 AAHCFRKYLDVSSWKVRAGSNILGNSPSLPVAKFIAEPNPLYPKPKDIALVKLQMPPLTF 300

301 SGTVRPICLPFFDEBELTPATPLWIGWFTKQNGKMSDILLQASVQVIDSTRCNADDAV 360
301 SGVVRPICLPFSDVLPATPVWVIGWFTTEENGKMSDMLQASVQVIDSTRCNADDAV 360

361 QGEVTEKMCAGIPGGVDTCCGSGGLMYTOSQWVHVGVISWGYCGGSPSTPGVYTKV 420
361 EGEVTEKMCAGIPGGVDTCCGSGGLMYTOSQWVHVGVISWGYCGGSPSTPGVYTKV 420

421 SAYLNIWNVKAEI 435
421 TAYLNIWNVKSEM 435

RESULT 3
Q8CFE0 PRELIMINARY; PRT; 471 AA.
AC Q8CFE0;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Similar to mosaic serine protease (Fragment).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
[1]
SEQUENCE FROM N.A.
RA STRAIN=C57BL/6J; TISSUE=Breast tumor;
RL Strausberg R.;
RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.
EMBL; BC042878; AAH42878.1; -.
KW Protease.
FT NON TER
SQ SEQUENCE 471 AA; 52535 MW; ED58CFE6B7C3BCC4 CRC64;

Query Match          30.1%; Score 704.5; DB 11; Length 471;
Best Local Similarity 36.2%; Pred. No. 8.3e-60;
Matches 153; Conservative 70; Mismatches 167; Indels 33; Gaps 9;

QY 25 ETRFKVGIPPIIALLSLASIIIVVLLIKVILDKYFFLCG-----OPLHPIPRKQL-CD 76
DB 65 ETRQKQ--LPLIGCVILLISLISLIL-----FYFWRGHTGIKYKPLESCPIHVRCD 116
```

```

QY 77 GBLDCLPLGDEEHCVKSFPEGPAPAVRLSKDRSTLOVLDSATGNWPSACFONFTEALAE 136
DB 117 GVVDCKMKSDELGC-----VRFWDKSLLLKVSYSSEWLPVCSWNWDTDSKR 165
QY 137 ACRQMGYSKPTFRAVEIGDQDLVDVEITENSQELMRNNSGFCILSGSLVSLHCLACGK 196
DB 166 TCQQLGFDG--AYRTTEVAHRDITSSFLLSYNTTIQESLYRSQCPSRRVSVLQCSHCG 223
QY 197 SLKTRVVGGEASVDSWPMQVSIQYDKQHVCGSILDPHWLTAHCF--RKHTDVFNW 254
DB 224 RAMTGRIVGALTSKPMQVSLHFGTHICGTLIDAQWLVTAHCFVTRKLEGW 283
QY 255 KVRAGSDKLGSPSLAVAKIIIEFNPMYKDN-DIALMKLQPLFSGTVRPICLPF 313
DB 284 KYVAGTSLNHLQPEASISQILINGNYTDEDDYDIALRLSKPLSLAHIPACLP 343
QY 314 BELTPATPLWIIGWFTKQNGKMSDILQASVQVVIDSTRCNADDAIOGEVTEKMCAGI 373
DB 344 QTFGLNETCWITGFGTKETDEKTSPLREVQVNLIDFKKNDLYVDSVLTFRMWCAGD 403
QY 374 PGGVDTCOGDSGGLM-YQSDQHVHVGVISWYGGCGSPSTGVTYKUSAYLNWIYVWK 432
DB 404 LRGGDSQCGSGGLPLVCQNNRWYLAGVTSWGTGCGQKPKGVTYKTVLPWIYRME 463
QY 433 AEL 435
DB 464 SEV 466

RESULT 4
Q9BYE2 PRELIMINARY; PRT; 581 AA.
ID AC Q9BYE2
AC Q9BYE2;
DT 01-JUN-2001 (TRENBLrel. 17, Created)
DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)
DT 01-WAR-2003 (TRENBLrel. 23, Last annotation update)
DE Membrane-type mosaic serine protease.
GN MSPL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1] SEQUENCE FROM N.A.
RC TISSUE=Lung;
RC MEDLINE=21167393; PubMed=11267681;
RA Kim D.R., Sharmin S., Inoue M., Kido H.;
RT "Cloning and expression of novel mosaic serine proteases with and
RT without a transmembrane domain from human lung.";
RL Biochim. Biophys. Acta 1518:204-209 (2001).
DR EMBL; AB048796; BAB39741.1; -.
DR HSSP; P00763; IDPO.
DR MEROPS; S01.087; -.
DR InterPro; IPR002172; LDL_receptor_A.
DR InterPro; IPR001254; Ser_protease_Try.
DR InterPro; IPR001190; Srcr_receptor.
DR Pfam; PF00057; ldl_recept_a; 1.
DR Pfam; PF00089; trypsin; 1.
DR SMART; SM00192; LDLa; 1.
DR SMART; SM00202; SR; 1.
DR SMART; SM00202; Tryp_Spc; 1.
DR PROSITE; PS0287; SRCR_2; 1.
DR PROSITE; PS0240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
DR Hydrolase; Protease; Serine protease.
DR SKW SEQUENCE 581 AA; 62689 MW; 4DABE24D7D5BA4A CRC64;

Query Match 29.7%; Score 695.5; DB 4; Length 581;
Best Local Similarity 35.4%; Pred. No. 8.4e-59;
Matches 152; Conservative 69; Mismatches 161; Indels 47; Gaps 11;

```

```

QY 22 IPWETFR--KVGIPILIALSLASIIIVVLLKVLDKYVFLCGQ-----PLH 67
DB 149 LKFTWREGQKPLIGCVLLIALVLSLIIL-----FOFWOGHTGIRHKQRESCPKH 202
QY 68 FIPRQLCDGELDCPLGDEEHCVKSFPEGPAPAVRLSKDRSTLOVLDSATGNWPSACFD 127
DB 203 AV-----RCDGVVDCKLKSDELGC-----VRFWDKSLLLKYSQSSQHWLPICSS 247
QY 128 NTEALAEATACRQMGYSKPTFRAVEIGDQDLVDVEITENSQELMRNNSGFCILSGSLV 187
DB 248 NWNDSYSEKTRCQLGFES--AHRTEVAHRDPFANGFSILRYNSTIQESLHRSHGCPQRYI 305
QY 188 SHHCLACGKSLKTPRVVGGEASVDSWPMQVSIQYDKQHVCGSILDPHWLTAHCF-- 245
DB 306 SLQCHSGCLRAMTGRIVGALASDSKPMQVSLHFGTHICGTLIDAQWLVTAHCFPV 365
QY 246 RKHTDVFNWVRAGSDKLGSPSLA-VAKIIIEFNPMY---PKNDIALMKLQPLF 301
DB 366 TREKVLGKWKVYAGTSLNHLQPEASIAEIIII--NSNYTDEDDYDIALMLSKPLTIS 422
QY 302 GTVRPICLPFDEELTPATPLWIIGWFTKQNGKMSDILQASVQVVIDSTRCNADDAIO 361
DB 423 AHIHPACLPMEHGTQPSLNETCWITGFGTKETDEKTSPLREVQVNLIDFKKNDLYVD 482
QY 362 GVTETKMCAGIPEGVDTCOGDSGGLM-YQSDQHVHVGVISWYGGCGSPSTGVTYK 420
DB 483 SYLTFRMWCAGDLHGGRDSQCGSGGLPLVCQNNRWYLAGVTSWGTGCGQKPKGVTYK 542
QY 421 SAYLNWIYN 429
DB 543 TEVLPIWYS 551

RESULT 5
Q8KITO PRELIMINARY; PRT; 453 AA.
ID AC Q8KITO
AC Q8KITO;
DT 01-OCT-2002 (TRENBLrel. 22, Created)
DT 01-OCT-2002 (TRENBLrel. 22, Last sequence update)
DT 01-WAR-2003 (TRENBLrel. 23, Last annotation update)
DE Type II transmembrane serine protease.
GN TMRPSS3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
[1] SEQUENCE FROM N.A.
RC STRAIN=clb-cj7;
RA Guipponi M., Scamuffa N., Scott H.S., Antonarakis S.E.;
RT "Isolation of the mouse Tmrpss3 genomic DNA sequence.";
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
DR EMBL; AJ429216; CAD2137.1; -.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR002172; LDL_receptor_A.
DR InterPro; IPR001254; Ser_protease_Try.
DR InterPro; IPR001190; Srcr_receptor.
DR Pfam; PF00057; ldl_recept_a; 1.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PF00722; CHYMOTRYPSIN.
DR SMART; SM00192; LDLa; 1.
DR SMART; SM00202; SR; 1.
DR SMART; SM00202; Tryp_Spc; 1.
DR PROSITE; PS01209; LDLRA_1; 1.
DR PROSITE; PS0068; LDLRA_2; 1.
DR PROSITE; PS0287; SRCR_2; 1.
DR PROSITE; PS0240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
DR Hydrolase; Protease; Serine protease; Transmembrane.
DR SKW SEQUENCE 453 AA; 49491 MW; 1ABCBF10AF61EF6 CRC64;

```



```

DR PROSITE; PS02040; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Protease; Serine protease.
SQ SEQUENCE 537 AA; 58102 MW; A39FF4E8816DAECF CRC64;

Query Match      29.3%; Score 686.5; DB 4; Length 537;
Best Local Similarity 38.6%; Pred. No. 5.7e-58;
Matches 142; Conservative 56; Mismatches 147; Indels 23; Gaps 7

QY   75 CDGELDCPLGEDDEHCVKSPPEGPAAVAVRLSKORSTLOVLDSATGNWFSAFDNPTFLAAL 134
Db   181 CDGVDPCKLSDSLGC-----VFEDWDKSLKIYGGSHQLWLPICSSNWINDSYS 229
QY   135 ETACRQMGYSKKTFERAIVEIGPDQDLVVHITENSQELMRNSSGPCLSGLSVLSIHLCLAC 194
Db   230 EKTCRQLGFES--AHRTTEVAHRDPAANSFIILRYNSTIOESLHRSCFSQRVI SLQCSHC 287
QY   195 GKSLKTPRVVVGGEASVDSPWPQWSIQYDKHQVCGGSILDPHWVLTAAHCF--RKHTDVDF 252
Db   288 GLRAMTGRIVGGALADSCKPWPQVSLHFQTHTCGGTLIDAQWVLTAAHCFPVTRKVL 347
QY   253 NIKVRAGSKLGSPSLA-VAKIIIEFNPMY---PKNDIALMKLOPFTFGSTVRPIC 308
Db   348 GWKYVAGTNSLHQLPEAASIAETII---NSNYTDEDDYDIAMRLSKPLTSAHIHPAC 404
QY   309 LPFFDELPAEPALWIIGWGFTKQNGKGKSDILLQASVOVIDSTRCNADDAYQGEVETKM 368
Db   405 LPMHGQTFSLNETCWITFGKTRHETDDKTSPPFLREYQVNLIIDFKCNDYLIVDSYLTPRM 464
QY   369 MCAGIEGGGYDTCCQSGSGPLM-YQSDQWHVGVISVMYGCGGPSTPGVTVKVSAYLNWI 427
Db   465 MCAGDLHGGRDSCQSGSGPLVCEQNRRWVLAVGTSMGTGGGRNKEGVTKVTEVLPWI 524
QY   428 YNVWKAEAL 435
Db   525 YSKWESEV 532

```

```

RESULT 7
Q9DGR2 PRELIMINARY; PRT; 767 AA.
ID Q9DGR2
AC Q9DGR2;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Embryonic serine protease-2.
OS XESP-2.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;
OC Xenopodinae; xenopus.
NCBI_TaxID=8355;
SEQUENCE FROM N.A.
RP MEDLINE=20363741; PubMed=10903452;
RX Yamada K., Takabatake T., Takeshima K.;
RT "Isolation and characterization of three novel serine protease genes
RT from Xenopus laevis.";
RL Gene 252:209-216(2000).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
DR EMBL; AB038497; BAB08217.1; -.
DR HSP; P00766; 1CHG.
DR MEROPS; S01.049; -.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR002172; LDL receptor A.
DR InterPro; IPR001254; Ser protease_Try.
DR Pfam; PF00057; ldl_recept_a; 4.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR PRINTS; PR00261; LDLRECEPTOR.
DR SMART; SM00192; LDLa; 8.
DR SMART; SM00020; Tryp SPG; 1.

```



```

81 LSCPGVSEKLLPSLK--AVSFRINGEDLLLEVQVRAPDMLLVCHEGWNPALGMHIC 138
139 ROMGYSSFTFRAVEIGDPDLDVVEITENSQELMRNS-----SGPCLSGSL 186
139 OSLGYFRLTQHKAVNL---SDIKL-----NRSQFAQLSARPGSLVEEAWQFSTNCPSGRI 191
187 VSLHCLACGKSLKTPRVVGGEEASVDSWPNQVSIQYDKQHVCGGSLDPHVVLTAAHC-- 244
192 VSLKCECGARPLASRIVGGQAVASGRWPQASVMLGSRHTCGGSLVAPYVWVTAACHMY 251
245 -FRKHTDVFNKVRAG---SDKLGFPPSLAVAKIIIBFNPMYPKN---DIALMKLQFP 297
252 SFRL-SRLSSWRVHAGLVSHSAVRQHGQTMVEKIIP---HPLYSAQNHDYDVALQLRTP 307
298 LTFSGTVRPICLPFDEELTPATPWIIGWFTKONGKMSDILLQASVQVIDSTRCNAD 357
308 INFSDTVSAVCLPAKEQHPQGSQGWGHTDPSHTSSDTLQDTMVPFLSTDLNCSS 367
358 DAYQGEVTEKMCAGIPGEGVDTCCGDSGGPLMYQS-DOWHVVGVSVGCGGSPSTPGV 416
368 CMYSGALTHRMCLCAGYLDGRADACQDSGGPLVCPSGDTWHLVGVVSWGRGCAENRPGV 427
417 YTKVSAVNLWYIN 429
428 YAKVAEFLDWDH 440

RESULT 12
Q8CDRO PRELIMINARY; PRT; 455 AA.
ID Q8CDRO;
AC Q8CDRO;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DE Adrenal mitochondrial protease short variant.
GN AMP.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NEDH;
RA Omer S., Bicknell A.B., Lowry P.J.;
RT "Identification of a rat adrenal mitochondrial protease.";
RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AFS37099; AAN06758.1; -.
KW Protease.
SQ SEQUENCE 455 AA; 49669 MW; B22EB2E7503C74B CRC64;

Query Match 27.0%; Score 631.5; DB 11; Length 455;
Best Local Similarity 36.2%; Pred. No. 1e-52;
Matches 135; Conservative 63; Mismatches 140; Indels 35; Gaps 10;

79 LDCPLGDEEHVKSFPFGPAVAVRLSKDRSTQLVDSATGNWFSACFDNFTALATAC 138
91 LNCPGVSRRELPLSLPK--TVSFRINGEDLLLVQVRAPDMLLVCHEGWNPALGMHIC 148
139 ROMGYSSKPTFRAVEIGDPDLDVVEITENSQELMRNS-----SGPCLSGSL 186
149 KSLGHRLTQHKAVNL---SDIKL-----NRSQFAQLSARPGSLVSWPVSANCPSGRI 201
187 VSLHCLACGKSLKTPRVVGGEEASVDSWPNQVSIQYDKQHVCGGSLDPHVVLTAAHC-- 244
202 VSLKCECGARPLASRIVGGQAVASGRWPQASVMLGSRHTCGGSLVAPYVWVTAACHMY 261
245 -FRKHTDVFNKVRAGSKDLGS---FSLAVAKIIIBFNPMYPKN---DIALMKLQFP 297
262 SFRL-SRLSSWRVHAGLVSHSAVRQHGQTMVEKIIP---HPLYSAQNHDYDVALQLRTP 317
298 LTFSGTVRPICLPFDEELTPATPWIIGWFTKONGKMSDILLQASVQVIDSTRCNAD 357

```

```

Db 318 INFSDTVSAVCLPAKEQHPQGSQGWGHTDPSHTSSDTLQDTMVPFLSTDLNCSS 377
QY 358 DAYQGEVTEKMCAGIPGEGVDTCCGDSGGPLMYQS-DOWHVVGVSVGCGGSPSTPGV 416
Db 378 CMYSGALTHRMCLCAGYLDGRADACQDSGGPLVCPSGDTWHLVGVVSWGRGCAENRPGV 437
QY 417 YTKVSAVNLWYIN 429
Db 438 YAKVAEFLDWDH 450

RESULT 13
Q8CJ16 PRELIMINARY; PRT; 371 AA.
ID Q8CJ16;
AC Q8CJ16;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DE Adrenal mitochondrial protease short variant.
GN AMP.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NEDH;
RA Omer S., Bicknell A.B., Lowry P.J.;
RT "Identification of a rat adrenal mitochondrial protease.";
RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AFS37099; AAN06758.1; -.
KW Protease.
SQ SEQUENCE 371 AA; 40694 MW; 89A64081D9A1FE26 CRC64;

Query Match 26.1%; Score 612; DB 11; Length 371;
Best Local Similarity 34.7%; Pred. No. 6.2e-51;
Matches 131; Conservative 62; Mismatches 138; Indels 46; Gaps 11;

QY 87 BEHCVSFPFGPA-----VAVRLSKDRSTQLVDSATGNWFSACFDNFTALTA 134
Db 2 EIRCTEGP-GPGIFRAELGDOQOQPIISFRINGEDLLLEVQVRAPDMLLVCHEGWNPALG 60
QY 135 ETACRMGYSSKPTFRAVEIGDPDLDVVEITENSQELMRNS-----SGPCL 182
Db 61 MHICQSLGYFRLTQHKAVNL---SDIKL-----NRSQFAQLSARPGSLVEEAWQFSTNCP 113
QY 183 SGLSVSHCLACGKSLKTPRVVGGEEASVDSWPNQVSIQYDKQHVCGGSLDPHVVLTAA 242
Db 114 SGRIVSLKCECGARPLASRIVGGQAVASGRWPQASVMLGSRHTCGGSLVAPYVWVTA 173
QY 243 HC---FRKHTDVFNKVRAG---SDKLGFPPSLAVAKIIIBFNPMYPKN---DIALMK 293
Db 174 HCMYSFRL-SRLSSWRVHAGLVSHSAVRQHGQTMVEKIIP---HPLYSAQNHDYDVALQL 229
QY 294 LQPLPFGTVRPICLPFDEELTPATPWIIGWFTKONGKMSDILLQASVQVIDSTR 353
Db 230 LRTPIFSDTVSAVCLPAKEQHPQGSQGWGHTDPSHTSSDTLQDTMVPFLSTDL 289
QY 354 CNADDAVQCEVTEKMCAGIPGEGVDTCCGDSGGPLMYQS-DOWHVVGVSVGCGGSP 412
Db 290 CNSSCMYSGALTHRMCLCAGYLDGRADACQDSGGPLVCPSGDTWHLVGVVSWGRGCAEN 349
QY 413 TPGVYTKVSAVNLWYIN 429
Db 350 RGVYAKVAEFLDWDH 366

RESULT 14
Q9DB10 PRELIMINARY; PRT; 799 AA.
ID Q9DB10;
AC Q9DB10;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)

```

01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE 1300008A22Rik protein (RIKEN cDNA 1300008A22 gene).
 TRPSS6 OR 1300008A22Rik.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NCBI_TaxID=10090;
 [1]
 SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Liver;
 RC MEDLINE=21085660; PubMed=11217851;
 RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaoka I.,
 RA Saito T., Okazaki Y., Gojobori T., Sono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
 RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 RA Nordone P., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
 RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
 RA Hayashizaki Y.;
 RA "Functional annotation of a full-length mouse cDNA collection."
 RL Nature 409:685-690(2001).
 [2]
 SEQUENCE FROM N.A.
 RC TISSUE=Liver;
 RC Strausberg E.;
 RA Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
 RL -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
 DR EMBL; AK004939; BAB23684.1; -;
 DR EMBL; BC029645; AAH29645.1; -;
 DR HSP; P00763; IDPO.
 DR MEROPS; S01.308; -;
 DR MGD; MGI:1919003; TmpRSS6.
 DR InterPro; IPR001314; Chymotrypsin.
 DR InterPro; IPR000859; CUB domain.
 DR InterPro; IPR002172; LDL receptor A.
 DR InterPro; IPR001254; Ser. protease_Try.
 DR Pfam; PF00057; ldl_recept_a; 2.
 DR Pfam; PF00089; tryptsin; 1.
 DR PRINTS; PR00722; CHYMOTRYPSIN.
 DR PRINTS; PR00261; LDLRECEPTOR.
 DR PROSITE; PS01180; CUB; 1.
 DR PROSITE; PS01209; LDLRA_1; 1.
 DR PROSITE; PS00068; LDLRA_2; 3.
 DR PROSITE; PS02040; TRYPSIN_DOM; 1.
 DR PROSITE; PS00134; TRYPSIN_HIS; 1.
 DR PROSITE; PS00135; TRYPSIN_SER; 1.
 DR HydroLase; Protease; Serine protease.
 SW SEQUENCE 799 AA; 89557 MW; 16315A646A4D5288 CRC64;
 Query Match 23.5%; Score 550.5; DB 11; Length 799;
 Best Local Similarity 33.2%; Pred. No. 1.9e-44;
 Matches 128; Conservative 60; Mismatches 146; Indels 51; Gaps 12;
 60 FLCQPLHFTPRKQLCDGELDCPLGDEEHCV-----KSPFEPGPAVAVRLSKDRSTLQVL 114
 450 FLCVNGLCVP---ACDGIKDCPNGLDERNCVCRAMPQCOEDSTCISLPRVCDRQP----- 502
 115 DSATGNWFSACFDNFTALAEACRQMGYSKPTFRAVEIGPQDLVDVVEITENSQELRM 174
 503 DCLNGSEEQCEQGVPCGTFCTFQCEDRSCVKKP-----NPECD----- 540
 175 RNSSGPGCLSGSLVSLHCLACGKSLKTPRVVGGEEASVDSWPQVSIQYDKQHVCGGSLD 234

Db 541 --QSDCRDGS--DEQHC--DGLQLSSRIVGTVSSEGEPMQASLQIRGRHICGGAIIA 596
 Qy 235 PHWLTAACFRKHT-----DVFNWKRAGSDKLGSPFSLAVAKIIIEFNPMYPKDN 287
 Db 597 DRWVITAACFCQDSDMASPKLMTVFLGKMRQNSRWPGEV-SFKVSRFLFLHPYHEEDSHDY 655
 Qy 288 DIALMKLQPLTFSTGVRPCLPFPDELTPTATPILWICGWFTKONGKMSDILLQASVQ 347
 Db 656 DVALQLDHPVVSIVRVPCLPFPDELTPTATPILWICGWFTKONGKMSDILLQASVQ 714
 Qy 348 VIDTRCNADDAVQGVYTKMKCAGIPEGGVTCQDGGGLMYQ--SDWHVGVIVSWG 405
 Db 715 LVFDLCS--EAVRYQVSPRLCAGYRKGKQACQDGGGLVCRFPSSGRWFLAGLVSWG 772
 Qy 406 YGGGSPSTGVTYKVSAYLNWYV 430
 Db 773 IGGGRNFFGVYTRVTRVINWIOQV 797

RESULT 15
 Q8CAN9 PRELIMINARY; PRT; 777 AA.
 ID Q8CAN9
 AC Q8CAN9
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Protease (Fragment).
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NCBI_TaxID=10090;
 [1]
 SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Thymus;
 RX MEDLINE=23354683; PubMed=12466851;
 RA THE FANTOM Consortium.
 RA the RIKEN Genome Exploration Research Group Phase I & II Team;
 RT "Analysis of the mouse transcriptome based on functional annotation of
 RT 60,770 full-length cDNAs."
 RL Nature 420:563-573(2002).
 DR EMBL; AK038356; BAC29973.1; -;
 FT NON TER 1
 SQ SEQUENCE 777 AA; 87314 MW; A18B2F4ECF06D3A8 CRC64;
 Query Match 23.3%; Score 545; DB 11; Length 777;
 Best Local Similarity 32.5%; Pred. No. 5.9e-44;
 Matches 124; Conservative 72; Mismatches 148; Indels 38; Gaps 13;
 69 IPRKQLCDGELDCPLGDEEHCVKSPFEPGPAVAVRLSKDRSTLQVL--SATGNWFSACPD 127
 410 IPLGNLDSYFCHDCDGSDEASCVRF-----LNGTSSNGLVQFNHSHIWLACAE 459
 128 NFEALAEACRQMGYSKPTFRAVEI---GPQDLVDVVEITENSQELRWNSSGPGCLSG 184
 460 NWTQISNEVCHLLGLGANSMPISSTGGP-----FVRVQAPNGSLILTFSLQCSQD 514
 185 SLVSIHC--LACGKSLKT---PRVVGEEASVDSWPQVSIQYDKQH-----VCGGSLD 234
 515 SLILLQCNHKSCEGKVTQKVPKIVGSGDAQAGAPWVVALYHRDRSTRLRLICGASLVS 574
 235 PHWLTAAC--FRKHTDVFNWKRAGSDKLGSPFSLAVAKIIIEFNPMYF---KDND 288
 575 SDNLVSAHCVYRNLDPTRWAVLGLHMQSNLTPQVVRVVDQIVINPHYDRRKVND 634
 289 IALMKLQPLTFSTGVRPCLPFPDELTPTATPILWICGWFTKONGKMSDILLQASVQ 348
 635 IAMHLEFKVNYTDYIQFICLPEENQIFIFPRTCSAGWGYDKINGSTV-DVLKEADVPL 693
 349 IDSTRCNADDAVQGVYTKMKCAGIPEGGVTCQDGGGLMYQ--SDWHVGVIVSWG 407
 694 ISNEKCO-QQLPEYNTITESMICAGYBEGGIDSCQDGGGLMCMQENNRNLFVGTSPGVQ 752

Tue Nov 25 11:44:32 2003

us-09-607-745-2.rspt

Page 9

408 CGGPSTPGVYTKVSAYLNWIYN 429
| | : | | | : | | : | | :
753 CALPNHPGVYVRVSQFIEWIHS 774

Search completed: November 23, 2003, 07:51:54
Job time : 71.2201 secs

GenCore version 5.1.16
Copyright (c) 1993 - 2003 CompuGen Ltd.

16M protein - nucleic search, using frame_plus_p2n model

Run on: November 23, 2003, 07:52:05 : Search time 5493.45 Seconds
(without alignments)
3239.440 Million cell updates/sec

Title: US-09-607-745-2

Perfect score: 2342

sequence: 1 MDPSDQPLNSLDVKPLRP.....VYTKSAYLNWYVWKAEL 435

coring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

MODEL=frame+ p2n.model -DEV=xlh
Q=/cgn2/1/USPRO.spool/US09607745/runat_21112003_144345_22184/app_query.fasta_1.1038
DB=GenEmbl -OFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
USPR=US09607745.cgn1_1_5308 @runat_21112003_144345_22184 -NCPU=3
NO_WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6
FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

atabase :

GenEmbl.*

1: gb.ba.*

2: gb.htg.*

3: gb.in.*

4: gb.om.*

5: gb.ov.*

6: gb.pat.*

7: gb.ph.*

8: gb.pl.*

9: gb.pr.*

10: gb.ro.*

11: gb.sts.*

12: gb.sy.*

13: gb.un.*

14: gb.vi.*

15: em.ba.*

16: em.fun.*

17: em.hum.*

18: em.in.*

19: em.mu.*

20: em.om.*

21: em.or.*

22: em.ov.*

23: em.pat.*

24: em.ph.*

25: em.pl.*

26: em.ro.*

27: em.sts.*

28: em.un.*

29: em.vi.*
30: em.htg.hum.*
31: em.htg.inv.*
32: em.htg.other.*
33: em.htg.mus.*
34: em.htg.pla.*
35: em.htg.rod.*
36: em.htg.man.*
37: em.htg.vrt.*
38: em.sy.*
39: em.htgo.hum.*
40: em.htgo.mus.*
41: em.htgo.other.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	2342	100.0	1305	6	AX076190 Sequence
2	2338	99.8	2038	6	AR142620 Sequence
3	2338	99.8	2038	6	BD137129 Human pro
4	2337	99.8	2081	9	AF179224 Homo sapi
5	2337	99.8	2088	9	BC011703 Homo sapi
6	2329	99.4	1479	6	AX076192 Sequence
7	2324	99.2	2079	6	AR232520 Sequence
8	2324	99.2	2079	6	AX207967 Sequence
9	2324	99.2	2079	9	AF216312 Homo sapi
10	2319	99.0	2137	6	AX207899 Sequence
11	2297.5	98.1	2063	6	AX092380 Sequence
12	2297.5	98.1	2063	6	AX376262 Sequence
13	2297.5	98.1	2063	6	AX395214 Sequence
14	2297.5	98.1	2063	6	AX697206 Sequence
15	2167	92.5	1793	9	BC012752 Homo sapi
16	1844	78.7	2060	10	BC021368 Mus muscu
17	1844	78.7	2068	10	AY043240 Mus muscu
18	1287	55.0	699	6	E13203 Human cDNA
19	704.5	30.1	2992	10	BC042878 Mus muscu
20	699.5	29.9	2067	6	AX149581 Sequence
21	697.5	29.8	1314	6	AX190543 Sequence
22	697.5	29.8	1314	6	AX190546 Sequence
23	697.5	29.8	1689	6	AX149579 Sequence
24	697.5	29.8	1748	6	AX369323 Sequence
25	694	29.6	609	6	AX076194 Sequence
26	692.5	29.6	2393	9	AB048796 Homo sapi
27	690.5	29.5	2237	9	AY190317 Homo sapi
28	687.5	29.4	3222	9	BC018715 Homo sapi
29	686.5	29.3	3507	5	AB038497 Xenopus l
30	683.5	29.2	1614	6	AX360089 Sequence
31	683.5	29.2	3324	9	AB048797 Homo sapi
32	683	29.2	2752	10	MMU300738 Mus muscu
33	677.5	28.9	2479	6	AR225372 Sequence
34	677.5	28.9	2479	6	AX201121 Sequence
35	677.5	28.9	2479	6	AX207965 Sequence
36	677.5	28.9	2479	6	AX267920 Sequence
37	677.5	28.9	2479	6	AX306771 Sequence
38	677.5	28.9	2479	6	AX395318 Sequence
39	677.5	28.9	2479	9	HSU75329 Human sarin
40	677.5	28.9	3966	6	AX041973 Sequence
41	676.5	28.9	1479	6	AR225345 Sequence
42	676.5	28.9	1738	6	AX395316 Sequence
43	676.5	28.9	1740	9	AF270487 Homo sapi
44	676.5	28.9	3195	9	BC051839 Homo sapi
45	675.5	28.8	1476	6	AX201158 Sequence

ALIGNMENTS

RESULT 1

AX076190
LOCUS AX076190 1305 bp DNA linear PAT 06-FEB-2001
DEFINITION Sequence 1 from Patent WO0104141.
ACCESSION AX076190
VERSION AX076190.1 GI:12710815
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE Suendermann, B., Hofmann, U., Matzku, S. and Wilbert, O.
JOURNAL Seripancrin
PUBLISHED Patent: WO 0104141-A 1 18-JAN-2001;
MERCK PATENT GmbH (DE)
FEATURES
source
1. .1305
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
1. .>1305
/notes="unnamed protein product"
/codon_start=1
/protein_id="CAC28479.1"
/db_xref="GI:12710815"
/translators="MOPSDOPSLDVKPLKPRIPMETPRKVGIPPIALLSLASI
IIVVLLKVLKDYFLCGPLHFI PKOLCDGLDCPLGDEBHCYKSPFEGPAVAV
RLSKDRSTLQVLSAANGNWSACFDNFTALAEATACRMGYSKPTFRFAVEIGPQDL
DVEITENSQELMRNNGSGLSIVSLVSLHCLAGSLKIPRVVGEERASVDWFWQV
SIQDXHGVGGSLIDPHWVLTAAHCFRHTDVFNVKRGSDKLGSPFSLAVAKII
IEFNPMYKNDIALMLKLPFLPTSGTVPICLPFFDELTATFLWIGWFTKQNG
GKMSDILLQASVVIDSTNCNADAYQGEVTERNMCAGIPEGGVDTTCQDSGGLMYQ
SDQHWVYIVSVGWYCGGPGTGVYTKVSNLYNWNKABL"
BASE COUNT 293 a 368 c 372 g 272 t
ORIGIN
Alignment Scores:
Pred. No.: 9.52e-215 Length: 1305
Score: 2342.00 Matches: 435
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
Gaps: 6
US-09-607-745-2 (1-435) x AX076190 (1-1305)
Y 1 MetAspProAspSerAspGlnProLeuAsnSerLeuAspVallylsProLeuArglyPro 20
b 1 ATGGATCTGACAGTGATCAACCTCTGAACAGCCTCGATGTCAAAACCCCTGGCAAAACC 60
Y 21 ArgileProMetGluThrPheArgLysValGlyIleProIleIleIleIleLeuLeuSer 40
b 61 CGTATCCCCATGGAGACCTTCAGAAAGGTGGGGATCCCATCATCATCATCATCATCTACTG 120
Y 41 LeuAlaSerIleIleIleValValValIleValIleValIleValIleValIleValIle 60
b 121 CTGGGAGATATCATCATCTGTGTCTCATCAAGGTGATCTCGGATAAATACTACTTC 180
Y 61 LeuCysGlyGlnProLeuHisPheIleProArgLysGlnLeuCysAspGlyGluLeuAsp 80
b 181 CTCGCGGGGACCTCTCCACTCATCCGAGGAGCAGCTGTGTGACGAGAGCTGGAC 240
Y 81 CysProLeuGlyGluAspGluHisCysVallylsSerPheProGluGlyProAlaVal 100
b 241 TGTCCCTTGGGGAGGACAGAGACACTGTGTCAAGAGCTTCCCGAGGCGCTGCAGTG 300
Y 101 AlaValArgLeuSerLysAspArgSerThrLeuGlnValLeuAspSerAlaThrGlyAsn 120
b 301 GCAGTCCGCTCTCCAGAGACCGATCCACATCGCAGGTGTGGACTCGGCCACAGGGAAC 360
Y 121 TrpPheSerAlaCysPheAspAsnPheThcGluAlaLeuAlaGluThrAlaCysArgGln 140
b 361 TGGTTCCTGTGCTGTTCGACAACTTCAGAAAGCTTCGCTGAGACAGCTGTAGGCAG 420

Qy 141 MetGlyTyrSerSerlylsProThrPheArgAlaValGluIleGlyProAspGlnAspLeu 160
Db 421 ATGGGCTACAGCAGCAAAACCCACTTTCAGAGCTGTGGAGATTGGCCACAGCAGATCTG 480
Qy 161 AspValValGluIleThrGluAsnSerGlnGluLeuArgMetArgAsnSerSerGlyPro 180
Db 481 GATGTTGTGAATACAGAAACACAGCAGGAGCTTCGATCGCGAACTCAAGTGGGCC 540
Qy 181 CysLeuSerGlySerLeuValSerLeuHisCysLeuAlaCysGlyLysSerLeuLysThr 200
Db 541 TGTCTCTCAGGCTCCCTGTCTCTCCCTGCACCTCTCTGCTGTGGAGAGAGCTGAAGACC 600
Qy 201 ProArgValValGlyGlyGluGluLaserValAspSerTrpProTrpGlnValSerIle 220
Db 601 CCCCGTGTGTGGTGGGAGAGAGGCTCTGTGGATTCTTGGCCCTTGGCAGGTCAAGATC 660
Qy 221 GlnTyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThr 240
Db 661 CAGTACAGCAACAGCAGCCTCTGTGGAGGAGCAGCATCTGGACCCCTCTGGCTCCTCAG 720
Qy 241 AlaAlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgLysSer 260
Db 721 GCAGCCCTCTGTCTCAGGAAACATACCCGATGTGTCACTGGAGGTGGCGCAGGCTCA 780
Qy 261 AspLysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleIleLeuPheAsn 280
Db 781 GACAACTGGGCAAGCTTCCCATCCCTGGCTGTGGCAAGATCATCATTCATTAATCAAC 840
Qy 281 ProMetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPhe 300
Db 841 CCATGTATACCCAAAGCAATGACATCGCCCTCATGAAGCTCAGTCCCATCTCATCTTC 900
Qy 301 SerGlyThrValArgProIleCysLeuProPheAspGluGluLeuThrProAlaThr 320
Db 901 TCAGGCACAGTCAGGCCCATCTGTCTGCCCTCTTGTGATGAGAGCTCACTCCAGCCACC 960
Qy 321 ProLeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyGlyLysMetSerAspIle 340
Db 961 CCATCTGGAGTATTTGGATGGGCTTTAGCAAGAGAGATGGAGGAGAGATGCTTGACATA 1020
Qy 341 LeuLeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAlaTyr 360
Db 1021 CTGCTGAGGCGTCAGTCCAGGTATTCAGCAGCACCGTGTCAATGCAGACCGATGCGTAC 1080
Qy 361 GlnGlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThr 380
Db 1081 CAGGGGAACTCACCGAGAGATGATGTGTGAGGATATCCCGAGAGGGGTGTGGACACC 1140
Qy 381 CysGlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGly 400
Db 1141 TGCAGGGTACAGTGGTGGGCCCCCTGATGTACCAATCTGACAGTGGCATGTGGTGGGC 1200
Qy 401 IleValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyValTrpThrLysVal 420
Db 1201 ATCGTTAGTGGGCTATGCTCGGGGGCCCCGAGCACCCGAGAGGTATACACCAAGGTC 1260
Qy 421 SerAlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu 435
Db 1261 TCAGCCTATCTCACTGGATCTACAATGTCTGAGGCGTGGCTG 1305

RESULT 2

AR142620

LOCUS

Sequence 18 from patent US 6203979.

ACCESSION

AR142620.1

VERSION

GI:15103906

KEYWORDS

Unknown.

ORGANISM

Unclassified.

REFERENCE

1. (bases 1 to 2038)

AUTHORS

Bandman, O., Hillman, J.L., Yue, H., Guegler, K.J., Corley, N.C.,

Tang, Y. Tom, and Shah, P.

Human protease molecules

Patent: US 6203979-A 18 20-MAR-2001;

JOURNAL

FEATURES

Location/Qualifiers

1. .2038

/organism="unknown"

BASE COUNT 462 a 591 c 569 g 416 t

ORIGIN

Alignment Scores:

Pred. No.: 4,11e-214 Length: 2038

Score: 2338.00 Matches: 434

Percent Similarity: 100.00% Conservatives: 1

Best Local Similarity: 99.77% Mismatches: 0

Query Match: 99.83% Indels: 0

DB: 6 Gaps: 0

US-09-607-745-2 (1-435) x AR142620 (1-2038)

Qy 1 MetAspProAspSerAspGlnProLeuAsnSerLeuAspVallyProLeuArgLysPro 20

Db 200 ATGGATCCTGACAGTATCAACCTCTGACAGCCTCGATGTCAACCCCTGGGCAACCC 259

Qy 21 ArgLeuProMetGluThrPheArgLysValGlyLeuProIleIleLeuLeuLeuSer 40

Db 260 CGTATCCCATGGAGACCTTCAGAAAGGTGGGGATCCCATCATATAGCAGTACTGAGC 319

Qy 41 LeuAlaSerIleIleLeuValValLeuLeuLeuValIleLeuAspLysTyrPhe 60

Db 320 CTGGGAGTATCATCTGTTGGTCTCTCAACAGGTGATCTGGATTAATACTACTTTC 379

Qy 61 LeuCysGlyGlnProLeuHisPheIleProArgLysGlnLeuCysAspGlyLeuLeuAsp 80

Db 380 CTCTGGGGGAGGCTCTCCACTTCATCCCGAGGAGCAGCTGTGTGACGAGAGTGGAC 439

Qy 81 CysProLeuGlyGlnAspGluGluHisCysValLysSerPheProGluGlyProAlaVal 100

Db 440 TGTCCTCTGGGGAGGAGCGAGGAGGAGTGTGTCAAGAGCTTCCCGAGGAGGCTGAGTG 499

Qy 101 AlaValArgLeuSerLysAspArgSerThrLeuGlnValLeuAspSerAlaThrGlyAsn 120

Db 500 GCAGTCCGCTCTCCAGGACCGATCCACTGCAGTGTGTGAGTCTGGCCACAGGGAAC 559

Qy 121 TrpPheSerAlaCysPheAspAsnPheThrGluAlaLeuAlaClnThrAlaCysArgGln 140

Db 560 TGGTCTCTCCCTGTTCGACACTTCACAGAGCTCTCGTAGAGACCTGTAGGAGCAG 619

Qy 141 MetGlyTyrSerLysProThrPheArgAlaValGluIleGlyProAspGlnAspLeu 160

Db 620 ATGGGCTACAGCAGCAAAACCCACTTTCAGAGCTGTGGAGATTGGCCACAGCAGATCTG 679

Qy 161 AspValValGluLeuThrGluAsnSerGlnGluLeuArgMetArgAsnSerSerGlyPro 180

Db 680 GATGTTGTTGAATATCAGAAACACCGAGAGCTTCGATCGGAGACTCAAGTGGGCCC 739

Qy 181 CysLeuSerGlySerLeuValSerLeuHisCysLeuAlaCysGlyLysSerLeuLysThr 200

Db 740 TGCTCTTCAGGCTCCCTGTCTCCCTGCACTGTCTTGGTGGGAGAGCCTGAAGACC 799

Qy 201 ProArgValValGlyGlyGluGluAlaSerValAspSerTrpProThrGlnValSerIle 220

Db 800 CCCCGTGTGGTGGGAGGAGGCTCTGTGGATTCGTGGCTTGGCAGTTCAGCATC 859

Qy 221 GlnTyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThr 240

Db 860 CAGTACGACAAACACGACGCTGTGGAGGAGGAGCATCTTGGACCCCTGCTGCTCAGC 919

Qy 241 AlaAlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySer 260

Db 920 GCAGCCCACTGCTCTCAGAAACATACCGATGTGTTCAACTGGAAGTGGGGCAGGCTCA 979

Qy 261 AspLysLeuGlySerPheProSerLeuAlaValAlaLysIleIleLeuLeuPheAsn 280

Db 980 GACAACTCGGCGAGCTTCCCATCCCTGGCTGGCCCAAGATCATCATCATGAATCAAC 1039

Qy 281 ProMetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPhe 300

Db 1040 CCATGTATCCCAAGACAAATGACATCGCCCTCATGAAGCTGCACTTCCCACTCACTTTC 1099

Qy 301 SerGlyThrValArgProIleCysLeuProPhePheAspGluGluLeuThrProAlaThr 320

Db 1100 TTAGGCGACAGTCAGGCCCATCTGTCTGCCCTTCTTGTAGAGAGCTCCTCCAGCCACC 1159

Qy 321 ProLeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyGlyLysMetSerAspIle 340

Db 1160 CCATCTCGATCATTTGATGGGCTTTTACGAAGCAGATGGAGGAGAGATGCTCTGACATA 1219

Qy 341 LeuLeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAlaTyr 360

Db 1220 CTGTCGAGCGCTCAGTCCAGTCAATGACAGCACGCGTGCATGACAGAGATGCGGTAC 1279

Qy 361 GlnGlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThr 380

Db 1280 CAGGGGGAAGTCACCGAGAAGATGATGTGTGAGCATCCCGAAGGGGTGTGGACACC 1339

Qy 381 CysGlnGlyAspSerClyGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGly 400

Db 1340 TGCAGGGGTGACAGTGTGGGCCCTGTGTACCAATCTGACCAAGTGGCATGTGTGGGC 1399

Qy 401 IleValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyValTyrThrLysVal 420

Db 1400 ATGTTAGCTGGGCTATGCTGGGGGCCCGACACCCCGAGGAGTATACCAAGGTC 1459

Qy 421 SerAlaTyrIleAsnTrpIleTyrAsnValTrpLysAlaGluLeu 435

Db 1460 TCAGCCTATCTCACTGGATCTACAAATGCTGGAAGGCTGAGCTG 1504

RESULT 3

BD137129

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

OS

PD

PF

PR

PI

PI

PI

PC

PC

CC

FH

FT

FT

FEATURES

source

Location/Qualifiers

1. .2038

/organism="Homo sapiens"

/mol_type="genomic DNA"

/db_xref="taxon:9606"

BASE COUNT 462 a 591 c 569 g 416 t

ORIGIN

Alignment Scores:

Pred. No.: 4,11e-214 Length: 2038
 Score: 2338.00 Matches: 434
 Percent Similarity: 100.00 Conservative: 1
 Best Local Similarity: 99.77 Mismatches: 0
 Query Match: 99.83 Indels: 0
 Gaps: 0

DB: 6
 US-09-607-745-2 (1-435) x BD137129 (1-2038)

QY 1 MetAspProAspSerAspGlnProLeuAsnSerLeuAspValLysProLeuArgLysPro 20
 DB 200 ATGGATCTGACAGTATCAACCTCTGACAGCCTCGATGTCAACCCCTCGGCAACCC 259
 QY 21 ArgIleProMetGluThrPheArgLysValGlyIleProIleIleIleAlaLeuLeuSer 40
 DB 260 CGTATCCCATGGAGACCTTCAGAAAGTGGGGATCCCATCATCATAGCACTACTGAGC 319
 QY 41 LeuAlaSerIleIleIleValValValLeuIleLysValIleLeuAspLysThrPhe 60
 DB 320 CTGGCGAGTATCATCATGTGTGTCTCATCAGAGTGTCTCGATTAATATCACTTC 379
 QY 61 LeuCysGlyGlnProLeuHisPheIleProArgLysGlnLeuAspGlyGluLeuAsp 80
 DB 380 CTCTGGGGCAGCCTCTCCACTTCATCCGAGAAAGCAGCTGTGTGACGAGAGCTGGAC 439
 QY 81 CysProLeuGlyGluAspGluGluHisCysValLysSerPheProGluGlyProAlaVal 100
 b 440 TGTCTCTGGGGGAGAGACGAGGAGCCTGTGTCAAGAGCTTCCCGAAGGCCCTGCGATG 499
 Y 101 AlaValArgLeuSerLysAspArgSerThrLeuGlnValLeuAspSerAlaThrGlyAsn 120
 b 500 CGAGTCCGGCTCTCCAGAGCCGATCCACACTGACAGTGTGGCTGGCTGGCCACAGGGAAC 559
 Y 121 TrpPheSerAlaCysPheAspSerPheThrGluAlaLeuAlaGluThrAlaCysArgGln 140
 b 560 TGGTCTCTGGCTGTTCACAACTTCACAGAGCTCTCGCTGAGACAGCTGTAGGCGAG 619
 Y 141 MetGlyThrSerLysProThrPheArgAlaValGluIleGlyProAspGlnAspLeu 160
 b 620 ATGGGCTACAGCAGCAAAACCCACTTCAGAGCTGTGGAGATGGCCACAGGATCTG 679
 Y 161 AspValValGluIleThrGluAsnSerGlnGluLeuArgMetArgAsnSerSerGlyPro 180
 b 680 GATGTGTGTAATCAGAAACACGAGGAGCTTCGCATCGCGAATCAAGTGGGCC 739
 Y 181 CysLeuSerGlySerLeuValSerLeuHisCysLeuAlaCysGlyLysSerLeuLysThr 200
 b 740 TGTCTCTAGGCTCCCTGTCTCCCTGCACCTGTCTGCTGTGGGAGAGCCTGAGACC 799
 Y 201 ProArgValValGlyGlyGluAlaSerValAspSerTrpProTrpGlnValSerIle 220
 b 800 CCCCCTGTGTGGTGGGAGAGGCTCTGTGGATTTCTTGGCTTGGCCTGGCAGTCCAGATC 859
 Y 221 GlnThrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThr 240
 b 860 CAGTACGACAAACAGACCTCTGTGGAGGAGCATCTTGGACCCCATCTGGTCTCTACG 919
 Y 241 AlaAlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySer 260
 b 920 GCAGGCCACTGTCTCAGGAAACATACCGATGTTCCTCACTGGAAGTGGGGCAGGCTCA 979
 Y 261 AspLysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleIleGluPheAsn 280
 b 980 GACAACTGGGAGCTTCCATCCCTGTGGTGGCCAGATCATCATGAAATTCAC 1039
 Y 281 ProMetTrpProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPhe 300
 b 1040 CCATGTACCCCAAGACATGACATCGCCCTCATGAGCTGCATCTCCACTCACTTTC 1099
 Y 301 SerGlyThrValArgProIleCysLeuProPhePheAspGluGluLeuThrProAlaThr 320

DB 1100 TCAGGCACGTTCAGGCCCATCTGTCTGCCCTTCTTTGATGAGGAGCTCACTCCAGCCACC 1159
 QY 321 ProLeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyGlyLysMetSerAspIle 340
 DB 1160 CCACCTCTGGATCATTTGGATGGGCTTTTACAGCAGAAATGGAGGGAAGATGTCTGACATA 1219
 QY 341 LeuLeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAlaTyr 360
 DB 1220 CTGCTGCAGGCTCAGTCCAGTCAATTGACAGCACACGGTGCATGACAGCATCGGTAC 1279
 QY 361 GlnGlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThr 380
 DB 1280 CAGGGGGAAGTCAACGAGAGATGATGTGTGAGGCATCCCGAAGGGGTGTGACACC 1339
 QY 381 CysGlnGlyAspSerGlyGlyProLeuMetTrpGlnSerAspGlnTrpHisValValGly 400
 DB 1340 TGCCAGGGTGACAGTGGTGGGCCCTGTATGATGATGATGATGATGATGATGATGATGAT 1399
 QY 401 IleValSerTrpGlyTrpGlyCysGlyGlyProSerThrProGlyValThrLysVal 420
 DB 1400 ATCGTTAGCTGGGCTATGGCTGGGGGCCCGAGCAGCAGGAGTATACACCAAGGTC 1459
 QY 421 SerAlaTyrLeuAsnTrpIleThrAsnValTrpLysAlaGluLeu 435
 DB 1460 TCAGCCTATCTCACTGGATCTACAATGTCTGGAAGGCTGAGCTG 1504

RESULT 4

AF179224 2081 bp mRNA linear PRI 08-JUN-2000
 LOCUS Homo sapiens transmembrane serine protease 3 (TMPRSS3) mRNA,
 DEFINITION complete cds.
 ACCESSION AF179224
 VERSION AF179224.1 GI:8347148
 KEYWORDS
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 REFERENCE 1 (bases 1 to 2081)
 AUTHORS Wallrapp,C., Hahnel,S., Muller-Pillasch,F., Burghardt,B.,
 Iwamura,T., Ruthenburger,M., Lerch,M.M., Adler,G. and Gress,T.M.
 TITLE A novel transmembrane serine protease (TMPRSS3) overexpressed in
 pancreatic cancer
 JOURNAL Cancer Res. 60 (10), 2602-2606 (2000)
 MEDLINE 20283276
 PUBMED 10825129
 REFERENCE 2 (bases 1 to 2081)
 AUTHORS Wallrapp,C. and Gress,T.M.
 TITLE Direct Submission
 JOURNAL Submitted (20-AUG-1999) Internal Medicine I, University of Ulm,
 Robert-Koch-Street 8, Ulm, Baden-Wuerttemberg 89081, Germany

FEATURES

source
 1..2081
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /chromosome="11"
 /map="11q23.3"
 /tissue_type="pancreatic carcinoma"
 1..2081
 /gene="TMPRSS3"
 215..1528
 /gene="TMPRSS3"
 /codon_start=1
 /product="transmembrane serine protease 3"
 /protein_id="AAF74526.1"
 /db_xref="GI:8347149"

gene

CDS

NGKXSDILLQASVQVVIDSTRNADDAAYQGEVTEKMMKAGIPBGVVDTCQDSDGGFLM
YQDQWVHVGVTSVMGYGCGSPSTPGVTKVSAYLNWIYNWKAEL"

misc_feature

215..307

/gene="TMPRS33"

/note="Region: cytoplasmic domain"

misc_feature

308..373

/gene="TMPRS33"

/note="Region: transmembrane domain"

misc_feature

374..1525

/gene="TMPRS33"

/note="Region: extracellular domain"

misc_feature

602..604

/gene="TMPRS33"

/note="glycosylation site"

misc_feature

746..748

/gene="TMPRS33"

/note="glycosylation site"

misc_feature

order(800..802,1142..1144)

/gene="TMPRS33"

/note="disulfide bond"

misc_feature

order(902..904,950..952)

/gene="TMPRS33"

/note="disulfide bond"

misc_feature

947..949

/gene="TMPRS33"

/note="active site"

misc_feature

1082..1084

/gene="TMPRS33"

/note="active site"

misc_feature

order(1280..1282,1328..1330)

/gene="TMPRS33"

/note="disulfide bond"

misc_feature

1373..1375

/gene="TMPRS33"

/note="active site"

BASE COUNT

484 a 597 c 576 g 424 t

RIGIN

Ligment Scores:

red. No.: 5.26e-214 Length: 2081

core: 2337.00 Matches: 434

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Very Match: 99.79% Indels: 0

B: 9 Gaps: 0

S-09-607-745-2 (1-435) x AF179224 (1-2081)

2 AspProAspSerAspGlnProLeuAsnSerLeuAspValLysProLeuArgLysProArg 21

224 GATCTGACAGTGATCAACCTCTGACAGCCTCGATGTCAAAACCCCTGGCCARACCCGT 283

22 IleProMetGluThrPheArgLysValGlyLeuProIleIleAlaLeuLeuSerIleu 41

284 ATCCCATCGGAGACCTTCAGAAAGTGGGATCCCAATCATAGCACTACTAGCCTG 343

42 AlaSerIleIleValValValLeuIleLysValIleLeuAspLysTyrTyrPheLeu 61

344 GCGAGTATCATCTGTGTCTCATCAGGATCTCGATGATCTGATTAATACTATTCTC 403

62 CysGlyGlnProLeuHisPheIleProArgLysGlnLeuCysAspGlyGluLeuAspCys 81

404 TGCAGGAGCCTCTCCACTTCATCCGAGGAGCAGCTGTGTGACGAGAGCTGGACTGT 463

82 ProLeuGlyGluAspGlnGluHisCysValLysSerPheProGluGlyProAlaValAla 101

464 CCCTTGGGGAGGAGCAGAGCAGCTGTGTCAAGAGCTTCCCGAAGGGCCTGCGAGTGCA 523

102 ValArgLeuSerLysAspArgSerThrLeuGlnValLeuAspSerAlaThrGlyAsnTrp 121

524 GTCCCGCCTCTCCAGACCGATCCACATCGAGGTGCTGGACTCGGCCACAGGAACTGG 583

122 PheSerAlaCysPheAspAsnPheThrGluAlaLeuAlaGluThrAlaCysArgGlnMet 141

Db 584 TTCTCTGCTGCTTTTCGACAACTTACAGAACTCTGCTGAGACAGCCTGTAGGCAGATG 643
QY 142 GlyTyrSerSerLysProThrPheArgAlaValGluIleGlyProAspGlnAspLeuAsp 161
Db 644 GGCTACAGCAGCAAAACCCACTTTACAGACTGTGAGATGTGGCCAGACCCAGATCTGAT 703
QY 162 ValValGluIleThrGluAsnSerGlnGluLeuArgMetArgAsnSerSerGlyProCys 181
Db 704 GTTGTGTAATACAGAAACACAGCCAGGAGCTTCGATGCGAACTCAAGTGGGCCCTGT 763
QY 182 LeuSerGlySerLeuValSerLeuHisCysLeuAlaCysGlyLysSerLeuLysThrPro 201
Db 764 CTCTCAGGCTCCCTGCTCTCCCTGCACTGTCTGCTGTGGAAAGAGCTGAAGACCCCC 823
QY 202 ArgValValGlyGlyGluAlaSerValAspSerTrpProTrpGlnValSerIleGln 221
Db 824 CGTGTGTGGTGGGAGGAGGCTCTGTGATTTCTGGCCTTGGCAGGTCAAGCATCCAG 883
QY 222 TyrAspLysGlnHisValCysGlySerIleLeuAspProHisTrpValLeuThrAla 241
Db 884 TAGCAGAAACAGCAGCTCTGTGGAGGAGCATCTCTGGACCCCACTGGGTCTCTCAGGCA 943
QY 242 AlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySerAsp 261
Db 944 GCCCACTGCTTCAGAAACATACCATGTGTTCACCTGAGAGGTGCGGGCAGGCTCAGAC 1003
QY 262 LysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleGluPheAsnPro 281
Db 1004 AAATGGGCGAGCTTCCCATCCCTGGCTGTGGCAAGATCATCATTTGAATTCACACCC 1063
QY 282 MetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPheSer 301
Db 1064 ATGTACCCCAAGACAATGACATCCCTCTCAAGAGCTGCAAGTTCCTCCACTCACTTTCTCA 1123
QY 302 GlyThrValArgProIleCysLeuProPhePheAspGluGluLeuThrProAlaThrPro 321
Db 1124 GGCACAGTCAGGCCCATCTGTCTGCCCTTCTTGTATGAGGAGCTCACTCCAGCCACCCCA 1183
QY 322 LeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyLysMetSerAspIleLeu 341
Db 1184 CTCTGGATCATTTGGATGGGCTTTACGAGCAGAGATGAGGAGGAGATGCTGACATACTG 1243
QY 342 LeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAlaThrGln 361
Db 1244 CTGCAGGCTCAGTCCAGTCAATGACAGCACCGTGCATGACAGCATGCGTACCGTACC 1303
QY 362 GlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThrCys 381
Db 1304 GGGGAAGTCACCGAGAGATGATGTGTGAGGATCCCGAAGGGGGTGTGGACACCTGC 1363
QY 382 GlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGlyIle 401
Db 1364 CAGGGTGACAGTGGTGGGCTTGTATGTACCAATCTGACCACTGGCATGTGTGGTGGCATC 1423
QY 402 ValSerTrpGlyTyrGlyCysGlyProSerThrProGlyValTyrThrLysValSer 421
Db 1424 GTTAGTGGGCTATGGCTGCGGGGCGGAGCACCCAGGAGTATACACCAAGTCTCA 1483
QY 422 AlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu 435
Db 1484 GCCTATCTCACTGGATCTACAATGTCTGGAAGGTGAGCTG 1525

RESULT 5

BC011703

LOCUS

DEFINITION

Homo sapiens, similar to mosaic serine protease, clone MGC:19490

IMAGE:3610695, mRNA, complete cds.

ACCESSION

BC011703

VERSION

MGC.

KEYWORDS

Homo sapiens (human)

SOURCE

ORGANISM

Homo sapiens

BC011703 2088 bp mRNA linear PRI 02-AUG-2001

Homo sapiens, similar to mosaic serine protease, clone MGC:19490

IMAGE:3610695, mRNA, complete cds.

ACCESSION BC011703

VERSION

MGC.

KEYWORDS

Homo sapiens (human)

SOURCE

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 2088)
Strausberg, R.
Direct Submission
Submitted (30-JUL-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgabbs-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC), Gaithersburg, Maryland;
Web site: <http://www.nisc.nih.gov/>
Contact: nisc.mgc@nih.gov
Shevchenko, Y., Wetherby, K.D., Beckstrom-Sternberg, S.M., Benjamin, B., Blakesley, R.W., Bouffard, G.G., Brinkley, C., Brooks, S., Dietrich, N.B., Guan, X., Gupta, J., Ho, S.-L., Karlins, E., Legaspi, R., Lim, M., Maduro, Q.B., Masiello, C., Mastrian, S.D., McCloskey, J.C., McDowell, J., Pearson, R., Snyder, B., Stantiripop, S., Thomas, P.J., Tionsgon, E.E., Touchman, J.W., Tsurgeon, C., Vogt, J.L., Walker, M.A., Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAL Plate: 27 Row: f Column: 19
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 6347148.

FEATURES

Location/Qualifiers
1..2088
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="MGC:19490 IMAGE:3610695"
/tissue_type="Pancreas, adenocarcinoma"
/clone_lib="NIH MGC 39"
/lab_host="DH10B-R"
/note="Vector: pOT87"
217..1530
/codon_start=1
/product="Similar to mosaic serine protease"
/protein_id="AAH11703.1"
/db_xref="GI:15079795"
/translation="MLQPPDSQPLNSLVKPLKRPIMETFRKVGIPILIALSLA
STIIIVVLIKVLKDYFLCGQPLHPIRKOLCDGELDCPLGDEBHCVKSPFPGPAV
AVRLSKDRSTLOVLPDSATGNWFSACFNFTALATAACROMGYSKPTFAVEIIGPQ
DLVDVETENSQELRNRSNGECLSGSLVSLHCLACGSLKTPRVVGGEEASVDSWPW
QVSIQVDKGVCHGCLDIPHWLTAACRKHDTFVNKKVAGSKDKGSPFSLAVAKI
IIIEENPMYPKNDIALMKLOPLFTSGVRPILCLFFDEELTPATPLWIIIGWGTFO
NGKSDIILQASVQIVDSTRNADDAQGEVTERMCAGIPEGVDTCCQSDSGPLM
YQSDQHWVGVISWGVCGGSTPGVYTKVAYLNWIVNWKAEI"
AGE COUNT 492 a 597 c 579 g 420 t
RIGIN

CDS

Alignment Scores:

Seq. No.: 5.28e-214 Length: 2088
Core: 2337.00 Matches: 434
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 99.79% Indels: 0
Gaps: 0

S-09-607-745-2 (1-435) x BC011703 (1-2088)

2 AspProAspSerAspGlnProLeuAsnSerLeuAspValLysProLeuArgLysProArg 21
|||||
226 GATCTTCAGATGATCACTCTGACACGCCCTGATGTCMAACCCCTCGCAACCCCGT 285

QY 22 IleProMetGluThrPheArgLysValGlyIleProIleIleAlaLeuLeuSerIleu 41
DB 286 ATCCCATCGAGACCTTCAGAAAGGTGGGATCCCATCATCATAGCACTACTGAGCCGTG 345
QY 42 AlaSerIleIleIleValValValLeuLeuIleValIleLeuAspLysTyrTyrPheLeu 61
DB 346 GCGAGTATCATCATTTGTTCTCTCATCAAGTGATCTCTGGATAAATACTACTTCTCCTC 405
QY 62 CysGlyGlnProLeuHisPheIleProArgLysGlnLeuCysAspGlyGluLeuAspCys 81
DB 406 TCGGGCAGCCTCTCCACTTCATCCCGAGGAGAGCTGTGTGACGAGAGCTGACTGT 465
QY 82 ProLeuGlyGluAspGluGlnHisCysValLysSerPheProGluGlyProAlaValAla 101
DB 466 CCGTTGGGGAGAGAGAGAGAGCACTGTGTACAGACTTCCCGAAGGGCCCTGACAGTGA 525
QY 102 ValArgLeuSerLysAspArgSerThrLeuGlnValLeuAspSerAlaThrGlyAsnTrp 121
DB 526 GTCCGCTCTCCAAAGAGCCGATCCACACTGCAGGTGTGTGGACTCGGCCACAGGAACTGG 585
QY 122 PheSerAlaCysPheAspAsnPheThrGluAlaLeuAlaGluThrAlaCysArgGlnMet 141
DB 586 TTCTCTGCTGCTGTTTGCACAACTTCACAGAGCTCTGCTGAGACAGCCTGTAGGCGAGTG 645
QY 142 GlyTyrSerSerLysProThrPheArgAlaValGluIleGlyProAspGlnAspLeuAsp 161
DB 646 GGCTACAGCAGCAAAACCCACTTTCAGAGCTGTGGAGATTGGCCACAGGACTCTGGAT 705
QY 162 ValValGluIleThrGluAsnSerGlnGluLeuArgMetArgAsnSerSerGlyProCys 181
DB 706 GTTGTGTAATCACAGAAACAGCAGGAGCTTCATCGCGNACTCAAGTGGGCCCCGT 765
QY 182 LeuSerGlySerLeuValSerLeuHisCysLeuAlaCysGlyLysSerLeuLysThrPro 201
DB 766 CTCTCAGGCTCCTGCTCTCCTCGCACGTCTTCTGCTGTGGAGATTGGCCACAGGCTGAGACCCCC 825
QY 202 ArgValValGlyGlyGluGluAlaSerValAspSerTrpProTrpGlnValSerIleGln 221
DB 826 CGTGTGCTGGTGGGAGAGAGGCTCTGTGGATTCTTGGCCTTGGCAGGTGAGCATCCAG 885
QY 222 TyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThrAla 241
DB 886 TAGCAAAACACAGCAGTCTGTGGAGGAGCATCTTGGACCCCACTGGGTCTCTCACGCA 945
QY 242 AlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySerAsp 261
DB 946 GCCCATCTGTTTCAGGAAACATACCGATGTGTTCATCTGAGAGGTGCGGGCAGGCTCAGAC 1005
QY 262 LysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleIleGluPheAsnPro 281
DB 1006 AAATGGGAGCTTCCCATCTTGGCTGTGGCCAGAGATCATCATTAATTCAACCCCTC 1065
QY 282 MetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPheSer 301
DB 1066 ATGTATCCCAAGAACAAATGACATCGCTCATGAAAGTGCAGTTCCTCCACTCACTTCTCA 1125
QY 302 GlyThrValArgProIleCysLeuProPheAspGluGluLeuThrProAlaThrPro 321
DB 1126 GGCAAGTCAGGCCCATCTGTCTGCCCTTCTTGTGATGAGGAGCTCACTCCAGCCACCCCA 1185
QY 322 LeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyGlyLysMetSerAspIleLeu 341
DB 1186 CTCTGATCATTTGGATGGGCTTTACGAGCAGAGATGAGGAGGAGATCTCTGACATCTG 1245
QY 342 LeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAlaTyrGln 361
DB 1246 CTGACGGCGCTAGTCCAGGTCTATTGACAGCACACGGTGCATATGACAGCAGTCCGTAC 1305
QY 362 GlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyValAspThrCys 381
DB 1306 GGGGAAGTCCCGAGAGAGATGTGTGAGGAGTCCCGGAGAGGGGGGTGTGACACCTGC 1365
QY 382 GlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGlyIle 401

```

Db      1366  CAGGGTGACAGTGGTGGGCCCCCTGATGATACCAATCTGACCACTGGCATGTGGTGGCAATC 1425
Qy      402  ValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyValThrIleValThrIleValSer 421
Db      1426  GTTAGCTGGGGCTATGCTGCGGGGGGCGGACACCCAGGAGTATACACCAAGGTCTCA 1485
Qy      422  AlaTyrLeuAsnTrpIleTyrAsnValTyrAsnValTyrAsnValTyrAsnValTyrAsnVal 435
Db      1486  GCCTATCTCACTGGATCTACCAATGCTGGAAGGCTGAGCTG 1527

RESULT 6
AX076192 LOCUS
DEFINITION Sequence 3 from Patent WO0104141.
ACCESSION AX076192
VERSION AX076192.1 GI:12710817
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE 1. Suendermann, B., Hofmann, U., Matzku, S. and Wilbert, O.
AUTHORS Seripancrin
TITLE Patent: WO 0104141-A 3 18-JAN-2001;
JOURNAL MERCK PATENT GmbH (DE)
FEATURES
source
1. 1479
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
1. 1479
/note="unnamed protein product"
/codon_start=1
/protein_id="CAC28480.1"
/db_xref="GI:12710818"
/translation="MDPDSQDLNLSLVKPKRPI PMETFRKVGIPILIIALLSLASI
IIVVLLIKVLDKIFLCGQPHFIPRKQDCDELDCPLGDEBHCVKSPFEGFANAV
RLSKDRSLQVLDLSTGNWFSACFDNFTAEATACRQMGYSRFTFRAVEIGPDOL
DYVEITENSQELRMNSGPCLSGLSLVSLHCLACSLKTRPVVGGEBASVDSWPQV
SIYDKQHVCGGSLILDPHWLTAARCFKRTDVFNNKVRAGSKLGSPPSLAVAKI1
IEFNPWPKNDIALMKQFPLTFSTGVRPICLPFDRELTAPTLIIGWFTKNG
GKMSILLQASVQVLDSTECNADDAQGEVEPKMWCAGIPSGVDVTCGDSGGFLMVQ
SQQWVGVLSVSGYCGGSESTGVITKYSALINWYINWKRDTIQRSNPSFTGLVIQ
QPAVPLMLRSARHTRUTTCVCFUSFPALQPLSLCPFPMD"

AGE COUNT 329 a 425 c 407 g 318 t
RIGIN

Alignent Scores:
red. No.: 1 97e-213 Length: 1479
core: 2329.00 Matches: 432
Percent Similarity: 100.00% Conservative: 0
est Local Similarity: 100.00% Mismatches: 0
Query Match: 99.44% Indels: 0
B: 6 Gaps: 0

S-09-607-745-2 (1-435) x AX076192 (1-1479)
Y 1 MetAspProAspSerAspGlnProLeuAsnSerLeuAspValLysProLeuArgLysPro 20
b 1 ATGGATCTCTGACAGTATCAACTCTGAACAGCCTCGATGTCAAAACCCCTGGGCAACCC 60
Y 21 ArgIleProMetGlnThrPheArgLysValGlyIleProIleIleIleAlaLeuLeuSer 40
b 61 CGPATCCCCATGAGACCTTCAGAAAGTGGGGATCCCCCATCATATGACACTACTGAGC 120
Y 41 LeuAlaSerIleIleIleValValValLeuIleLysValIleLeuAspLysTyrTrpPhe 60
b 121 CTGGCGAGTATCATCATTTGGTGTCTTCATCAGAGTGTCTTGATTAATATCTACTTC 180
Y 61 LeuCysGlyGlnProLeuHisPheIleProArgLysGlnLeuLeuAspGlyLeuLeuAsp 80

```

```

Db      181  CTCTGGGGCAGCCTCTCCACTTCTCCGAGGAGCAGCTGTGTACGGAGAGCTGGAC 240
Qy      81  CysProLeuGlyGlnAspGlnGluHisCysValLysSerPheProGluGlyProAlaVal 100
Db      241  TGTCCCTTGGGGGAGGACGAGGAGCACTGTGTCAAGAGCTTCCCCGAAGGCTTGCAAGT 300
Qy      101  AlaValArgLeuSerLysAspArgSerThrLeuGlnValLeuAspSerAlaThrGlyAsn 120
Db      301  GCAGTCCGGCTCTCCAAGGACCGATCCACACTGCAAGTCTGCACTCGGCCACAGGAAAC 360
Qy      121  TrpPheSerAlaCysPheAspAsnPheThrGluAlaLeuAlaGluThrAlaCysArgGln 140
Db      361  TGTCTTCTGCTGCTGTGTTCACAACTTCACAGAGCTCTCCCTTGAGACAGCTGTAGGCG 420
Qy      141  MetGlyTyrSerSerLysProThrPheArgAlaValGluIleGlyProAspGlnAspLeu 160
Db      421  ATGGGCTCAGCAGCAAAACCCACTTTCAGAGCTGTGGAGATTGGCCCAACAGGATCTG 480
Qy      161  AspValValGluIleThrGluAsnSerGlnGluLeuArgMetArgAsnSerSerGlyPro 180
Db      481  GATGTGTGTGAATATCAGAAAACAGCCAGGAGCTTCGCATCGGAACTCAAGTGGGCC 540
Qy      181  CysLeuSerGlySerLeuValSerLeuHisCysLeuAlaCysGlyLysSerLeuLysThr 200
Db      541  TGTCTCTCAGGCTCCCTGCTCTCCCTGCACTGTCTTGGCTGTGGAGAGAGCTTGAAGAC 600
Qy      201  ProArgValValGlyGlyGluAlaSerValAspSerTrpProTrpGlnValSerIle 220
Db      601  CCCGCTGTGTGGTGGGAGGAGGCTCTGTGTGATTTCTTGGCTTGGCAGGTCAACATC 660
Qy      221  GlnTyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThr 240
Db      661  CAGTACGACAAACAGCACGCTCTGTGGAGGAGCCTCTGACACCCCACTGGTCTCTCACG 720
Qy      241  AlaAlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySer 260
Db      721  GCAGCCCACTGCTTCAGGAAACATACCGATGTCTTCAACTGGAGGTGGCGGAGCTCA 780
Qy      261  AspLysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleIleGluPheAsn 280
Db      781  GACAAACTGGCGAGCTTCCCATCCCTGGCTGTGGCCCAAGATCATCATCTGAATCAAC 840
Qy      281  ProMetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPhe 300
Db      841  CCCATGTACCCCAAGACATGATGATCGCCCTCATGAAGCTGCAGTCCCATCTCATCTTC 900
Qy      301  SerGlyThrValArgProIleCysLeuProPhePheAspGluLeuThrProAlaThr 320
Db      901  TCAGGCACAGTCAGGCCCATCTGTCTGCCCTCTTTGTATGAGAGCTCACTCCAGCCACC 960
Qy      321  ProLeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyGlyLysMetSerAspIle 340
Db      961  CCACTCTGGATCATTTGGATGGGGCTTTACGAAAGAGATGGAGGGAAGATGTCTGACATA 1020
Qy      341  LeuLeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAspAlaTyr 360
Db      1021  CTGCTGCAGCGCTCAGTCCAGGCTCATTTGACGACACACGCTGCAATGCACGATGGTAC 1080
Qy      361  GlnGlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyValAspThr 380
Db      1081  CAGGGGGAATTCACCGAGAGAGATGATGTGTGAGGCGCATCCCGAAAGGGGTGTGACACC 1140
Qy      381  CysGlnCysSerGlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGly 400
Db      1141  TGCAGGGGTGACAGTGGTGGGCCCCCTGTGTGATGTACCAATCTGACAGTGGCATGTGG 1200
Qy      401  IleValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyValThrLysVal 420
Db      1201  ATCGTTAGCTGGGCTATAGCTGTGGGGGGCCCGAGACACCCGAGGATATACCAAGGTC 1260
Qy      421  SerAlaTyrLeuAsnTrpIleTyrAsnValTrpLys 432
Db      1261  TCAGCCTATCTCACTGGATCTACAAATGTCTGAAG 1296

```

```

RESULT 7
AR232520
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE
1 (bases 1 to 2079)
Mack, D., Gish, K.C. and Wilson, K.E.
Methods of diagnosing colorectal cancer, compositions, and methods
of screening for colorectal cancer modulators
Patent: US 645668-A 1 24-SEP-2002;
Location/Qualifiers
1..2079
/organism="unknown"
BASE COUNT 489 a 594 c 575 g 421 t
ORIGIN

Alignment Scores:
Pred. No.: 9,26e-213 Length: 2079
Score: 2324.00 Matches: 434
Percent Similarity: 99.77% Conservative: 0
Best Local Similarity: 99.77% Mismatches: 0
Query Match: 99.23% Indels: 1
DB: 6 Gaps: 0

US-09-607-745-2 (1-435) x AR232520 (1-2079)
Y 2 AspProAspSerAspGlnProLeuAsnSerLeuAspValLysProLeuArgLysProArg 21
b 217 GATCCTGACAGTGATCAACCTCTGACAGCCTCGATGTCAACCCCTGGCGAACCCCGT 276
Y 22 IleProMetGluThrPheArgLys-ValGlyLeuProIleIleIleAlaLeuSerIle 41
b 277 ATCCCCATGAGACCTTCAGAAAGTGGGGATGCCCATCATAGCAGTACTAGAGCT 336
Y 41 uAlaSerIleIleIleValValValLeuIleLysValIleLeuAspLysTyrPheIle 61
b 337 GCGAGTATCATCTTGTGTGTCTCATCAGGTGATCTCGATTAATACTACTTCT 396
Y 61 uCysGlyGlnProLeuHisPheIleProArgLysGlnLeuCysAspGlyGluLeuAspCy 81
b 397 CTCGGGCGAGCCTCTCCACTTCAATCCCGAGGAGCAGCTGTGTGACGGAGCTGAGCTG 456
Y 81 sProLeuGlyGluAspGluGluHisCysValLysSerPheProGluGlyProAlaValAl 101
b 457 TCCCTTGGGGGAGGAGAGGAGCAGTGTGTACAGACTTCCCGAGGAGCCTGAGTGGC 516
Y 101 aValArgLeuSerLysAspArgSerThrLeuGlnValLeuAspSerAlaThrGlyAsnTr 121
b 517 AGTCCGCTCTCCAGAGCCGATCCACACTGACAGGTGCTGAGCTCGGCCACAGGGAATG 576
Y 121 pPheSerAlaCysPheAspAsnPheThrGluAlaLeuAlaGluThrAlaCysArgGlnMe 141
b 577 GTTCTCTGCTGTTTTCACAACTTCACAGAGCTCTCGCTGACAGAGCTGTAGGAGAT 636
Y 141 tGlyTyrSerSerLysProThrPheArgAlaValGluIleGlyProAspGlnAspLeuAs 161
b 637 GGGCTACAGCAGCAACCCACTTTCAGAGCTGTGAGATGTGCCACAGCAGGATCTGA 696
Y 161 pValValGluIleThrGluAsnSerGlnGluLeuArgMetArgAsnSerSerGlyProCy 181
b 697 TGTGTGTAATACAGAAACAGCCAGGAGCTTCGCATGCGGAATCAAGTGGGCGCTG 756
Y 181 sLeuSerGlySerLeuValSerLeuHisCysLeuAlaCysGlyLysSerLeuLysThrPr 201
b 757 TCTCTAGGCTCCCTGCTCTCCCTGCACTGTCTGTGCTGTGGGAGAGGCTGAAGACCCC 816
Y 201 oArgValValGlyGlyGluGluAlaSerValAspSerTrpProTrpGlnValSerIleGl 221

```

```

Db 817 CCCTGTGGTGGGTGGGAGAGGCTCTGTGTGATCTTGGCCTTGGCAGGTGAGCATCA 876
Y 221 nTyrAspLysGlnHisValCysGlySerIleLeuAspProHisTrpValLeuThrAl 241
Db 877 GTACGACAAACAGCAGCTGTGTGGAGGAGCATCTCTGGACCCCACTGGGTCTCACGCG 936
Y 241 alaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySerAs 261
Db 937 AGCCCACTCTTCAGGAAACATACCGATGTGTTCACAGGAGGTGCGGCGAGGCTCAGA 996
Y 261 pLysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleGluPheAsnPr 281
Db 997 CAACTGGGCGACTTCCCATCTCTGGCTGTGGCCCAAGATCATCATTAATTAACACC 1056
Y 281 oMetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPheSe 301
Db 1057 CATGTATCCCAAGACATGACATGCGCTCATGAGCTGCAAGTCCCACTCACCTTTCTC 1116
Y 301 rGlyThrValArgProIleCysLeuProPheAspGluGluLeuThrProAlaThrPr 321
Db 1117 AGGCACAGTCAGGCCCATCTGTCTGCCCTTCTTGTATGAGGAGCTCACTCCAGCCACCC 1176
Y 321 oLeuTrpIleIleGlyTyrPheThrLysGlnAsnGlyLysMetSerAspIleLeu 341
Db 1177 ACTCTGGATCATTTGGATGGGCTTTACGAAAGCAGAAATGGAGGGAAGATGCTTCGACATACT 1236
Y 341 uLeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAspAlaTyrGl 361
Db 1237 GCTCGAGGCGTCACTCAGGTCAATGACAGCACAGTGCATGAGCATGCGATGCGTACCA 1296
Y 361 nGlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThrCy 381
Db 1297 GGGGGAAGTCACCGAAGATGATGTGACAGCATCCCGAAGGGGTGTGGACACCTG 1356
Y 381 sGlnGlyAspSerGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGlyLvl 401
Db 1357 CCAGGTGACAGTGGGGCCCTGATGTACCAATCTGACCACTGACAGTGGCATGTGGTGGGCT 1416
Y 401 eValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyValTyrThrLysValSe 421
Db 1417 CGTTAGCTGGGCTATGGCTGGGGGGCCGAGCAGCCAGGAGTATACCAAGGTCTC 1476
Y 421 rAlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu 435
Db 1477 AGCTATCTCACTGGATCTACATGTCTGGAAGGCTGAGCTG 1519

RESULT 8
AX207967 2079 bp DNA linear PAT 31-AUG-2001
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1
AUTHORS
Maddison, B.L., Ong, E.O. and Yeh, J.C.
Nucleic acid molecules encoding transmembrane serine proteases, the
encoded proteins and methods based thereon
Patent: WO 0157194-A 71 09-AUG-2001;
CORVAS INTERNATIONAL, INC. (US)
FEATURES
Location/Qualifiers
1..2079
/organism="Homo sapiens"
/mol type="genomic DNA"
/db xref="taxon:9606"
251..1522
/note="Nucleotide sequence encoding transmembrane
protease, serine 4 (TPRSS4)"
codon_start=1
CDS

```

```

/protein id="CAC60389.1"
/db_xref="GI:15422564"
/translation="MSNPANFVSPWPSVSGVPIIALLSLASIIIVVLKIVLD
KYIFLGGQPLHPIFRKQUCDELDCPLGDEDEHCVKSPFEGPAVAVRLSKDRSTLQVL
DSATGNWFSACDFNFTALACROMGYSKPTFRVFIQDQDLVDVEITENSQBL
RMNNSGFCLSGLSVLSLACAGSLKPTFRVVGEEASVDSPQWQVGIQDKQHVCGG
SILDPKWLTAHGFRTKHTDVFNNKVRAGSKLGSFSLAVAKIIITEFNMVFPKND
IALMKLOPLTFSGTVRPICLPFPEDELTPTPLWIIIGWFTKONGKNSDIILOASV
OVIDSTRCANADAYCGEYTERMKMAGIEGGVDICQSDSGPLMYQSDQHWVGVISW
GYGCGSPSTPGYIKVSAYLNIWYNWKAEI"
BASE COUNT      489 a   594 c   575 g   421 t
ORIGIN

```

```

Alignment Scores:
Pred. No.:      9,26e-213      Length:      2079
Core:           2324.00        Matches:      434
Percent Similarity: 99.77%      Conservative: 0
Best Local Similarity: 99.77%      Mismatches: 0
Query Match:      99.23%      Indels:      1
DB:               6           Gaps:        0

```

JS-09-607-745-2 (1-435) x AX207967 (1-2079)

```

QY 2 AspProAspSerAspGlnProLeuAsnSerLeuAspVallysProLeuArgysProArg 21
DB 217 GATCCTGACAGTATCAACCTCTGACAGCCTCGATGTCAAAACCCCTCGCAACCCCGT 276
QY 22 IleProMetGluThrPheArgLys-ValGlyIleProIleIleIleLeuSerLe 41
DB 277 ATCCCATGGAGACCTTCAGAAAGTGTGGGATCCCATCATCATAGACACTACTGAGCCT 336
QY 41 uLaSerIleIleIleValValLeuIleLysValIleLeuAspLysTyrTyrPheLe 61
b 337 GCGCAGTATCATCATGTGTCTCATCAAGGTGATTCGGATAAATACTACTTCTCT 396
Y 61 uCysGlnProLeuHisPheIleProArgLysGlnLeuCysAspGlyGluLeuAspCy 81
b 397 CTGCGGCGAGCTCTCCACTTCATCCGAGGAGCAGCTGTGTGACGAGAGCTGAGCTG 456
Y 81 sProLeuGlyGluAspGluGluHisCysValLysSerPheProGluGlyProAlaValAl 101
b 457 TCCCTTGGGGAGGAGGAGGAGCAGCTGTGTCAAGAGCTTCCCGAAGGCGCTGAGTGGC 516
Y 101 aValArgLeuSerLysAspArgSerThrLeuGlnValLeuAspSerAlaThrGlyAsnTr 121
b 517 AGTCGCGCTCTCCAAGGACCGATCCACACTGCAGTGTCTGGACTCGGCCACAGGAACTG 576
Y 121 pPheSerAlaCysPheAspAsnPheThrGluAlaLeuAlaGluThrAlaCysArgGlnMe 141
b 577 GTTCTCTGCTCTTTCGACAACTTCACAGAGCTCTGCTGAGACAGCTGTAGGCAGAT 636
Y 141 tGlyTyrSerSerLysProThrPheArgAlaValGluIleGlyProAspGlnAspLeuAs 161
b 637 GGGCTACAGCAGCAACCCACTTTCAGAGCTGTGGAGATTGGCCAGACCAAGGATCTGGA 696
Y 161 pValValGluIleThrGluAsnSerGlnGluLeuArgMetArgAsnSerSerGlyProCy 181
b 697 TGTGTGTAANTACAGAAAACAGCAGGAGCTTGCATGCGGAATCAAGTGGGCCCTG 756
Y 181 sLeuSerGlySerLeuValSerLeuHisCysLeuAlaCysGlyLysSerLeuLysThrPr 201
b 757 TCTCTCAGGCTCCCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 816
Y 201 oArgValValGlyGlyGluGluAlaSerValAspSerTrpProTrpGlnValSerIleGl 221
b 817 CCGTGTGGTGGTGGGAGAGGAGGCTCTGTGATTTCTGGCTTGGCAGGTGAGCATCCA 876
Y 221 nTyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThrAl 241
b 877 GTACGACACACAGCAGCTCTGTGGAGGAGCATCTGGNACCCCTCTGGTCTCTCAGGCG 936
Y 241 alaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySerAs 261

```

```

Db 937 AGCCCACTGCTTCAGGAACATACCGATGTGTCAACTGGAAGGTGCGGCGAGGCTCAGA 996
QY 261 physLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleGluPheAsnPr 281
Db 997 CAAACTGGGCGACTCCCATCCCTGGCTGTGCCAAGATCATCATCAATGAATCAACCC 1056
QY 281 oMetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPheSe 301
Db 1057 CATGTACCCCAAGACATGACATCGCCCTCATGAAGCTGCAGTTCCTCCACTCATTTCTC 1116
QY 301 rGlyThrValArgProIleCysLeuProPheAspGluGluLeuThrProAlaThrPr 321
Db 1117 AGSCACAGTCAGGCCCATCTGTCTGCCCTTCTTTGATGAGGAGCTCACTCCAGCCACCCC 1176
QY 321 oLeuTyrIleIleGlyTyrGlyPheThrLysGlnAsnGlyGlyLysMetSerAspIleLe 341
Db 1177 ACTCTGATCATTTGGATGGGGCTTTACGAGACAGATGAGGAGGAGAGATGCTTGACATACT 1236
QY 341 uLeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAspAlaTyrGl 361
Db 1237 GCTGCAGGCGTCACTCCAGTCAATTCACAGCACACGGTGCATGCGAGCATGCGGTACCA 1296
QY 361 nGlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThrCy 381
Db 1297 GGGGAGGATCACCGAGAGATGATGTGCGAGCATCCCGAAGGGGTGTGGACACTG 1356
QY 381 sGlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGlyI 401
Db 1357 CCAGGTGACAGTGTGTGGCCCTGATGTACAACTGACACAGTGGCATGTGTGGGCT 1416
QY 401 eValSerTrpGlyTyrGlyCysGlyProSerThrProGlyValTyrThrLysValSe 421
Db 1417 CGTACCTGGGCTATGGCTGCGGGCGCCGAGCACCCAGAGATATACCAAGGTCTC 1476
QY 421 xAlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu 435
Db 1477 AGCTATCTCACTGGATCTACAACTCTGGAAGGTGAGCTG 1519

```

RESULT 9

AF216312

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

CDS

1..2079

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

251..1522

/note="MT-SP2"

/codon_start=1

/product="type II membrane serine protease"

/protein_id="AAF31436.1"

/db_xref="GI:6911219"

/translation="MSNPANFVSPWPSVSGVPIIALLSLASIIIVVLKIVLD

KYIFLGGQPLHPIFRKQUCDELDCPLGDEDEHCVKSPFEGPAVAVRLSKDRSTLQVL

2079 bp mRNA linear PRI 07-FEB-2000
Homo sapiens type II membrane serine protease mRNA, complete cds.

Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 2079)
Smeekens, S.S., Lorimer, D.D., Wang, E., Hou, J. and Linnevers, C.
MT-SP2, a novel type II membrane serine protease expressed in
trachea, colon, and small intestine: identification, cloning, and
chromosomal localization
Unpublished

2 (bases 1 to 2079)
Smeekens, S.S., Lorimer, D.D., Wang, E., Hou, J. and Linnevers, C.
Direct Submission
Submitted (14-DEC-1999) Axy's Pharmaceuticals, Inc, 180 Kimball Way,
South San Francisco, CA 94080, USA
Location/Qualifiers

1..2079
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
251..1522
/note="MT-SP2"
/codon_start=1
/product="type II membrane serine protease"
/protein_id="AAF31436.1"

/translation="MSNPANFVSPWPSVSGVPIIALLSLASIIIVVLKIVLD
KYIFLGGQPLHPIFRKQUCDELDCPLGDEDEHCVKSPFEGPAVAVRLSKDRSTLQVL

DSATGNWFSACFDNTEALAEATACRQMGYSKPTPRAVEIGPDODLDWEITENSQEL
 RNRSSGCLSLVSLHCLACGLKTPRVVGGESVDSNPQVWSIYDKQHVCGG
 IALMLQFLPTSGTVRPTCLPFFDEELTPATPLMIIGWFTKONGKMSDILQASV
 QVIDSTRCNADDAVOEVTETKMMKAGIPEGGVDTCCQSDSGPPLMYQSDQWHEVGVSW
 GYGGGSPSPFGVYTKVSAVLNWIYNWKAEL"
 BASE COUNT 489 a 594 c 575 g 421 t
 ORIGIN

Alignment Scores:
 Pred. No.: 9,26e-213 Length: 2079
 Score: 2324.00 Matches: 434
 Percent Similarity: 99.77% Conservative: 0
 Best Local Similarity: 99.77% Mismatches: 0
 Query Match: 99.23% Indels: 1
 DB: 9 Gaps: 0

JS-09-607-745-2 (1-435) x AF216312 (1-2079)
 QY 2 AspProAspSerAspGlnProLeuAsnSerLeuAspValLysProLeuArgLysProArg 21
 DB 217 GATCTCGAGTGAATCACTCTGAACACAGCTCGATGTCAACCCCTGCGCAACCCCGT 276
 QY 22 IleProMetGluThrPheArgLys-ValGlyLeuProIleIleIleAlaLeuLeuSerIle 41
 DB 277 ATCCCATGGAGACCTTCAGAAAGTGTGGGATCCCATCATCATAGCACTACTGAGCCT 336
 QY 41 uAlaSerIleIleValValValValLeuLeuLeuValIleLeuAspLysTyrTyrPheIle 61
 DB 337 GCGAGTATCATATTGTGGTGTCTCATCAAGGTGATTCGGATAAATACTATTCTCT 396
 QY 61 uCysGlyGlnProLeuHisPheIleProArgLysGlnLeuCysAspGlyGluLeuAspCy 81
 DB 397 CTCCGGGCGACCTCTCCACTTCATCCGAGGAAGAGCTGTGTGCGGAGAGCTGCACTG 456
 QY 81 sProLeuGlyGluAsnGluLysCysValLysSerPheProGluGlyProAlaValAl 101
 DB 457 TCCCTTTGGGGAGGAGGAGAGGACACTGTGTCAAGAGCTTCCCGAAGGCGCTGCACTG 516
 QY 101 aValArgLeuSerLysAspArgSerThrLeuGlnValLeuAspSerAlaThrGlyAsnTr 121
 b 517 AGTCCGCTCTCAAGACCGATCCACACTGCAAGTGTCTGCACTCGGCCACAGGACCTG 576
 QY 121 pPheSerAlaCysPheAspAsnPheThrGluAlaLeuAlaGluThrAlaCysArgGlnMe 141
 DB 577 GTTCTCTGCTGTTCGACAACTTCAGAAAGCTCTCGCTGAGACAGCCTGTAGGACAGT 636
 QY 141 tGlyTyrSerSerLysProThrPheArgAlaValGluLeuGlyProAspGlnAspLeuAs 161
 DB 637 GGGCTACAGACAGCAACCCACTTTCAGAGCTGTGGAGATTGCCACAGACAGGATCTGA 696
 QY 161 pValValGluIleThrGluAsnSerGlnGluLeuArgMetArgAsnSerSerGlyProCy 181
 b 697 TGTGTGTGAATCACAAGAAACAGCCAGGAGCTTCGCATGCGGAATCAAGTGGGCGCTG 756
 Y 181 sLeuSerGlySerLeuValSerLeuHisCysLeuAlaCysGlyLysSerLeuLysThrPr 201
 b 757 TCTCTCAGGCTCCTGGTCTCCCTGCACTGTCTGTGCTGTGGAGAGGCTTGAAGACCC 816
 Y 201 oArgValValGlyGlyGluGluAlaSerValAspSerTrpProTrpGlnValSerIleGl 221
 b 817 CCGTGTGTGGTGGGAGGAGGAGGCTCTGTGGATTCTTGGCCTTGGCAGGTGAGCATCCA 876
 Y 221 nTyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThrAl 241
 b 877 GTACGCAAAACAGACCTCTGTGTGGAGGAGCATCTCTGGACCCCACTGGGTCTCTCAG 936
 Y 241 aAlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySerAs 261
 b 937 AGCCCACTGCTTCAGGAAACATACCGATGTGTTCACTGGAAGGTCTCGGCGAGGCTCAGA 996
 Y 261 pLysLeuGlySerPheProSerIleAlaValAlaLysIleIleIleIleGluPheAsnPr 281

DB 997 CAAACTGGGCGAGCTTCCCATCCCTGGCTGGCCAGATCATCATCATGATTCAACCC 1056
 QY 281 oMetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPheSe 301
 DB 1057 CATGTACCCCAAGACAATGACATGCCCTCATGAAGCTGCAAGTCTCCCACTCACTTTCTC 1116
 QY 301 xGlyThrValArgProIleCysLeuProPhePheAspGluGluLeuThrProAlaThrPr 321
 DB 1117 AGGCACAGTCAGCCCATCTGTCTGCCCTTCTTTGATGAGGAGCTCATCTCCAGCCACCC 1176
 QY 321 oLeuTrpIleIleGlyTyrGlyPheThrLysGlnAsnGlyGlyLysMetSerAspIleIle 341
 DB 1177 ACTCTGCATCATGTGATGGGCTTTTACGAAGCAGAAATGGAGGAAAGATGCTCTGACAT 1236
 QY 341 uLeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAspAlaTyrGl 361
 DB 1237 GCTCCAGGCGTCAGTCCAGTCAATGACAGCACAGGTGCAATGACAGCATGCGTATCCA 1296
 QY 361 nGlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThrCy 381
 DB 1297 GGGGGAAGTCACCGAGAGATGATGTGTGACAGCATCCCGGAAGGGGTGTGACACCTG 1356
 QY 381 sGlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGlyL 401
 DB 1357 CCAGGCTGACAGTGTGGGCGCTGTGTGTACCAATCTGACCAAGTGTGTGTGGGCAT 1416
 QY 401 eValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyValTyrThrLysValSe 421
 DB 1417 CGTTAGCTGGGCTATGCTGTGGGGGCGGAGCACCACCGAGGATATACACCAAGGTCTC 1476
 QY 421 rAlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu 435
 DB 1477 AGCCTATCTCACTGGATCTACAATGTCTGGAAGGCTGAGCTG 1519
 RESULT 10
 AX207899 2137 bp DNA linear PAT 31-AUG-2001
 LOCUS Sequence 3 from Patent WO0157194.
 DEFINITION AX207899
 ACCESSION AX207899
 VERSION AX207899.1 GI:15422497
 KEYWORDS
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Rukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1
 AUTHORS Madison, E.L., Ong, E.O. and Yeh, J.C.
 TITLE Nucleic acid molecules encoding transmembrane serine proteases, the
 encoded proteins and methods based thereon
 JOURNAL Patent: WO 0157194-A 3 09-AUG-2001;
 CORVAS INTERNATIONAL, INC. (US)
 FEATURES
 source Location/Qualifiers
 1. .2137
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 261..1574
 /note="DNA sequence encoding a transmembrane serine
 protease (MSP3) protein"
 /codon_start=1
 /protein_id="CAC60378.1"
 /db_xref="GI:15422498"
 /translation="MTQDPDSQPLNSLDVLPKPRIPMETFRKVGPIIALLSLA
 SIIIVVLKYLIDKYFLGQELHPIPKQLCDGELDCLGDEHCHVKSPPEGPV
 AVLSKDRSTLQVDSATGNWFSACFDNTEALAEATACRQMGYSKPTPRAVEIGPD
 DLVDVITENSQELNRSSGCLSLVSLHCLACGLKTPRVVGGESVDSNPQVWSIYDKQ
 IIEFNPMYPKNDIADHMLQFLPTSGTVRPTCLPFFDEELTPATPLMIIGWFTKON
 NGKMSDILQASVOVIDSTRCNADDAVOEVTETKMMKAGIPEGGVDTCCQSDSGPPL
 YQSDQWHEVGVSWGYGGGSPSPFGVYTKVSAVLNWIYNWKAEL"
 508 a 612 c 589 g 428 t
 BASE COUNT
 ORIGIN

Alignment Scores:

Pred. No.: 2 89e-212 Length: 2137
 Core: 2319.00 Matches: 432
 Percent Similarity: 99.54% Conservative: 0
 Best Local Similarity: 99.54% Mismatches: 2
 Query Match: 99.02% Indels: 0
 DB: 6 Gaps: 0

US-09-607-745-2 (1-435) x AX207899 (1-2137)

2 AspProAspSerAspGlnProLeuAsnSerLeuAspValLysProLeuArgLysProArg 21
 270 GATCCTGACAGTATCACTTCAACAGCTCGATGTCACACCCCTCGCAACCCCGT 329
 22 IleProMetGluThrPheArgLysValGlyLeuProLlellelleAlaLeuSerLeu 41
 330 ATCCCCATGGAGACCTTCAGAAAGTGGGGATCCCATCATCATAGCACTACTGAGCCTG 389
 42 AlaSerIlellelleValValValLeuIleLysValIleLeuAspLysTyrTyrPheLeu 61
 390 GCGAGTATCATCTGTTGTCTCATCAAGGTGATTCGGATATATATCTTTCTTC 449
 62 CysGlyGlnProLeuHisPheIleProArgLysGlnLeuCysAspGlyGluLeuAspCys 81
 450 TGGGGGAGCCTCTCCACTTCATCCGAGGAGCAGCTGTGTGACGGAGAGCTGGACTGT 509
 82 ProLeuGlyGluAspGluGluHisCysValLysSerPheProGluGlyProAlaValAla 101
 510 CCGTTGGGGAGGACGAGGACATGTGTCAAGAGCTTCCCGAGAGGCCCTGCGAGTGCA 569
 102 ValArgLeuSerLysAspArgSerThrLeuGlnValLeuAspSerAlaThrGlyAsnTrp 121
 570 GTCCGCTCTCCAGGACCGATCCACACTGACAGTGTGTGACTGGCCACAGGAACTGG 629
 122 PheSerAlaCysPheAspAsnPheThrGluAlaLeuAlaGluThrAlaCysArgGlnMet 141
 630 TTCTCTGCTGTTTCGACAACTTCACAGAGCTCTGCTGAGACAGCTGTGAGCAGATG 689
 142 GlyTyrSerLysProThrPheArgAlaValGluIleGlyProAspGlnAspLeuAsp 161
 590 GGCTACAGCAGCAACCCACTTCAGAGCTGTGAGATTTGCCCCAGACCAAGATCTGGAT 749
 162 ValValGluIleThrGluAsnSerGlnGluLeuArgMetArgAsnSerSerGlyProCys 181
 750 GTTGTGAATTCAGAAACAGCAGAGCTTCGATCGGAATCAAGTGGGCCCTGT 809
 182 LeuSerGlySerLeuValSerLeuHisCysLeuAlaCysGlyLysSerLeuLysThrPro 201
 810 CTCTCAGGCTCCCTGGTCTCCCTGCACTGTCTTGCCTGTGGGAAGAGCTGAAGACCCC 869
 202 ArgValValGlyGlyGluGluAlaSerValAspSerTrpProTrpGlnValSerIleGln 221
 870 CGTGTGTGGTGGGAGGAGGCTCTGTGATTTCTGGCTTGGCAGGTGAGCATCCAG 929
 222 TyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThrAla 241
 930 TAGCACAATACAGCAGCTCTGTGAGGAGGAGCATCTCTGACCCCACTGGTGTCTCACGGCA 989
 242 AlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySerAsp 261
 990 GCCCACTGCTTCAGAAACATACCGATGTGTTCAACTGGAAGGTGGGGCAGGCTCAGAC 1049
 262 LysLeuGlySerPheProSerLeuAlaValAlaLysIlellellellelleGluPheAsnPro 281
 1050 AACTGGGAGCTTCCCATCCCTGGCTGTGGCCAGATCATCATCATGTAATTCACACCC 1109
 282 MetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPheSer 301
 1110 ATGTACCCAAAGACATGATCGCCCTCATGAAGTGGAGTTCAGTTCCTCATCTTCTCA 1169
 302 GlyThrValArgProLleCysLeuProPhePheAspGluGluLeuThrProAlaThrPro 321
 1170 GGCACAGTCAGGCTCATCTGTCTGCCCTTTCTTGTGAGGAGGCTCACTCCAGCCACCCCA 1229

QY 322 LeuTrpIlelleGlyTrpGlyPheThrLysGlnAsnGlyGlyLysMetSerAspIleLeu 341
 Db 1230 CTCTGGATCATTTGGATGGGCTTTACGAAGCAGATGGAGGAAGATCTCTGACATCTG 1289
 QY 342 LeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAspAlaTyrGln 361
 Db 1290 CTGCAGCGCTCAGTCCAGGTCAATTGACAGCACACGGTGCATGACAGACGATCGGTACCAG 1349
 QY 362 GlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThrCys 381
 Db 1350 GGGGAATCACCAGAGATGATGTGTGAGGATCTCCCGAGAGGGGTGTGGACACCTGC 1409
 QY 382 GlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGlyIle 401
 Db 1410 CAGGGTACACAGTGTGGGCCCCCTGATGTACCAATCTGACAGTGGCATGTGGTGGGCATC 1469
 QY 402 ValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyValTyrThrLysValSer 421
 Db 1470 GTTAGCTGGGGCTATGGCTCGGGGGCCCCGAGCACCCGAGAGTATACCAAGGTCTCA 1529
 QY 422 AlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu 435
 Db 1530 GCCTATCTCACTGGATCTACAAATGCTGAGAGCTGAGCTG 1571

RESULT 11
 AX092380 2063 bp DNA linear PAT 21-MAR-2001
 LOCUS Sequence 111 from Patent WO0116318.
 DEFINITION AX092380
 ACCESSION AX092380.1 GI:13444504
 VERSION
 KEYWORDS
 SOURCE Homo sapiens (human)
 ORGANISM
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1
 AUTHORS Eaton,D.L., Filvaroff,E., Gerritsen,M.E., Goddard,A.,
 Godowski,P.J., Grimaldi,C.J., Gurney,A.L., Watanabe,C.K. and
 Wood,W.I.
 TITLE Secreted and transmembrane polypeptides and nucleic acids encoding
 the same
 JOURNAL Patent: WO 0116318-A 111 08-MAR-2001;
 Genentech, Inc. (US)
 FEATURES
 source
 1..2063
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /db_xref="taxon:9606"
 BASE COUNT 477 a 591 c 576 g 419 t
 ORIGIN

Alignment Scores:
 Pred. No.: 3.17e-210 Length: 2063
 Score: 2297.50 Matches: 429
 Percent Similarity: 98.85% Conservative: 0
 Best Local Similarity: 98.85% Mismatches: 0
 Query Match: 98.10% Indels: 5
 DB: 6 Gaps: 1

US-09-607-745-2 (1-435) x AX092380 (1-2063)

QY 2 AspProAspSerAspGlnProLeuAsnSerLeuAspValLysProLeuArgLysProArg 21
 Db 219 GATCCTGACAGTATCAACCTCTGACAGCTCGATGTCAACCCCTCGCAACCCCGT 278
 QY 22 IleProMetGluThrPheArgLysValGlyIleProLlellellelleAlaLeuSerLeu 41
 Db 279 ATCCCATGGAGACCTTCAGAAAGTGGGGATCCCATCATCATAGCACTACTGAGCCTG 338
 QY 42 AlaSerIlellellelleValValValLeuIleLysValIleLeuAspLysTyrTyrPheLeu 61
 Db 339 GCGAGTATCATCTGTTGTGTCTTCATCAAGGTGATTCGTGATAAATACTACTTCTTC 398

282 MetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPheSer 301
1044 ATGTACCCCAAGACAAATGATCGCCCTCATGAAGCTGCAGTTCCCACTCTTCTCA 1103
302 GlyThrValArgProIleCysLeuProPhePheAspGluGluLeuThrProAlaThrPro 321
1104 GGCACAGTCAGGCCCATCTGTCTGCCCTTCTTTGATGAGGAGCTCACTCCAGCACCCTCA 1163
322 LeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyGlyLysMetSerAspIleLeu 341
1164 CTCTGGATCATTCGATGGGCTTTACGAGCAGAAATGGAGGAGATGCTCGACATCTG 1223
342 LeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAlaThrGln 361
1224 CTGCAGCGCTCAGTCCAGGTCATTCAGACGACACGGTGCATGACAGCATGCTACACG 1283
362 GlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThrCys 381
1284 GGGGAAGTCCACGAGAGATGATGTGTGAGCATCCGGAGGGGTGTGGACACTGC 1343
382 GlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGlyIle 401
1344 CAGGCTGACAGTGTGTGGCCCTCATGTACCAATCTGACCATGTCATGTGGTGGCATC 1403
402 ValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyValTyrThrLysValSer 421
1404 GTTAGCTGGGGCTATGGCTGGGGGGCCGAGCACCAGGAGTATACACCAAGGTCTCA 1463
422 AlaTyrLeuAsnTrpIleTyrAsnValTyrLysAlaGluLeu 435
1464 GCCTATCTCACTGGATCTACATGCTGGAGGCTGAGCTG 1505

RESULT 14
X697206
OCUS
EPINITION
SEQUENCE 274 from Patent WO0078961.
X697206
X697206.1 GI:29498146
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Ferrara, N., Stewart, T.A., Williams, P.M., Baker, K.P., Desnoyers, L.,
Eaton, D.L., Gao, W.Q., Pan, J., Botstein, D., Fong, S., Goddard, A.,
Godowski, P.J., Gurrey, A.L., Smith, V., Tumas, D., Wood, W.I.,
Grimaldi, C.J., Hillan, K.J., Paoni, N.F., Roy, M.A., and Watanabe, C.K.
Secreted and transmembrane polypeptides and nucleic acids encoding
the same
Patent: WO 0078961-A 274 28-DEC-2000;
Genentech Inc. (US)
Location/Qualifiers
1..2063
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
477 a 591 c 576 g 419 t
ASE COUNT 477 a 591 c 576 g 419 t
RIGIN

Alignment Scores:
red. No.: 3.17e-210 Length: 2063
core: 2297.50 Matches: 429
Percent Similarity: 98.85% Conservative: 0
Best Local Similarity: 98.85% Mismatches: 0
Query Match: 98.10% Indels: 5
B: 6 Gaps: 1

US-09-607-745-2 (1-435) x AX697206 (1-2063)
2 AspProAspSerAspGlnProLeuAsnSerLeuAspValLysProLeuArgLysProArg 21

Db 219 GATCTTGACAGTGAATCAACTCTGAACAGCCTCGATGTCTAAACCCCTGCGCAAAACCCCGT 278
QY 22 IleProMetGluThrPheArgLysValGlyIleProIleIleIleAlaLeuLeuSerLeu 41
Db 279 ATCCCATCGAGACCTTCAGAAAGTGGGATCCCATCATCATGACACTACTGAGCCTG 338
QY 42 AlaSerIleIleIleValValValLeuIleLysValIleLeuAspLysTyrTyrPheLeu 61
Db 339 GCAGATATCATCTTGTGTGTCTCATCAAGGTGATTTCTGATAAATACTACTTCTCTC 398
QY 62 CysGlyGlnProLeuHisPheIleProArgLysGlnLeuCysAspGlyGluLeuAspCys 81
Db 399 TGGGCGAGCCTCTCCACTTCATCCCGAGAGACAGTGTGACGAGAGCTGGACTGT 458
QY 82 ProLeuGlyGluAspGluGluHisCysValLysSerPheProGluGlyProAlaValAla 101
Db 459 CCCTTGGGGGAGGAGGAGGACACTGTGTCAAGAGCTCTCCCGAAGGCTGTCAGTGGCA 518
QY 102 ValArgLeuSerLysAspArgSerThrLeuGlnValLeuAspSerAlaThrGlyAsnTrp 121
Db 519 GTCCGCTCTCCAAAGACCGATCCACACTGCAGTGTGACTCGGCCACAGGGAACCTGG 578
QY 122 PheSerAlaCysPheAspAsnPheThrGluAlaLeuAlaGluThrAlaCysArgGlnMet 141
Db 579 TTCTCTGCTGTTCGACAACTTCAGAAAGCTCTCGTGAGACAGCTGTAGGCGATG 638
QY 142 GlyTyrSerSerLysProThrPheArgAlaValGluIleGlyProAspGlnAspLeuAsp 161
Db 639 GGCTACAGC-----AGAGCTGTGGAGATTGGCCACACAGGATCTGGAT 693
QY 162 ValValGluIleThrGluAsnSerGlnGluLeuArgMetArgAsnSerSerGlyProCys 181
Db 684 GTTGTGAAATCACAGAAACAGCCAGGAGCTTCGCATCGGAACCTCAAGTGGGCCCTGT 743
QY 182 LeuSerGlySerLeuValSerLeuHisCysLeuAlaCysGlyLysSerLeuLysThrPro 201
Db 744 CTCTCAGGCTCCCTGTCTCCCTGCACTGTCTTGCCTGTGGAGAGCCTGAAAGACCCCC 803
QY 202 ArgValValGlyGlyGluAlaSerValAspSerTrpProTrpGlnValSerIleGln 221
Db 804 CGTGTGTGGTGGGAGGAGGAGGCTCTGTGGATTCTTGGCCTTGGCAGTTCAGATCCAG 863
QY 222 TyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThrAla 241
Db 864 TAGCAAAACAGCACCTCTCTGTGAGGAGGAGCATCTCGAAGGAGGAGGAGGAGGAGGAG 923
QY 242 AlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySerAsp 261
Db 924 GCCCACTGCTTCAGGAAACATACCGATGTTCCTGAAAGTGGGAGGAGGAGGAGGAGGAG 983
QY 262 LysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleIleGluPheAsnPro 281
Db 984 AAACCTGGGAGCTTCCATCCCTGGCTGTGGCCAGATCATCATCATTTGAATTCACCCCC 1043
QY 282 MetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPheSer 301
Db 1044 ATGTACCCCAAGACAAATGACATCGCCCTCATGAAGCTGCAGTTCCCACTCTCTCA 1103
QY 302 GlyThrValArgProIleCysLeuProPhePheAspGluGluLeuThrProAlaThrPro 321
Db 1104 GGCACAGTCAGGCCCATCTGTCTGCCCTTCTTTGATGAGGAGCTCACTCCAGCACCCTCA 1163
QY 322 LeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyGlyLysMetSerAspIleLeu 341
Db 1164 CTCTGGATCATTCGATGGGCTTTACGAGCAGAAATGGAGGAGATGCTCGACATCTG 1223
QY 342 LeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAlaThrGln 361
Db 1224 CTGCAGCGCTCAGTCCAGGTCATTCAGACGACACGGTGCATGACAGCATGCTACACG 1283
QY 362 GlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThrCys 381
Db 1284 GGGGAAGTCCACGAGAGATGATGTGTGAGCATCCGGAGGGGTGTGGACACTGC 1343

382 GlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGlyIle 401
 1344 CAGGGTGACAGTGGTGGCCCTGATGATACCAATCTGACCAAGTGGATGGTGGGCATC 1403
 402 ValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyValTyrThrIysValSer 421
 1404 GTTAGCTGGGGCTATGCTGCGGGGGCCGAGACCCAGAGGTATACCAAGGTTCTCA 1463
 422 AlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu 435
 1464 GCTATCTCACTGAGTCTACATGCTGGAAGCTGAGCTG 1505

RESULT 15

BC012752 1793 bp mRNA linear PRI 20-AUG-2001
 Homo sapiens, similar to transmembrane protease, serine 4, clone
 IMAGE:3623466, RNA, partial cds.

BC012752
 BC012752.1 GI:15215322

KEYWORDS

ORGANISM

Source

Homo sapiens (human)

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 1793)

Strausberg,R.

Direct Submission

Submitted (15-AUG-2001) National Institutes of Health, Mammalian

Gene Collection (MGC), Cancer Genomics Office, National Cancer

Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,

USA

NIH-MGC Project URL: <http://mgc.nci.nih.gov>

Contact: MGC help desk

Email: cgabbs-remail.nih.gov

Tissue Procurement: ATCC

CDNA Library Preparation: Rubin Laboratory

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LML)

DNA Sequencing by: Institute for Systems Biology

<http://www.systemsbio.org>

contact: amadan@systemsbiology.org

Anup Madan, Rachel Dickhoff, Jessica Fahey, Stephanie Ford, Julia

Greene, Mark Kettman and Anuradha Madan

Clone distribution: MGC clone distribution information can be found

through the I.M.A.G.E. Consortium/LML at: <http://image.llnl.gov>

Series: IRAL Plate: 24 Row: m Column: 19.

Location/Qualifiers

1..1793

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:3623466"

/tissue_type="Pancreas, adenocarcinoma"

/clone_lib="NIH_MGC_39"

/lab_host="DH10B-R"

/notes="vector: pOTB"

<1..1219

/codon_start=2

/product="Similar to transmembrane protease, serine 4"

/protein_id="AAH12752.1"

/db_xref="GI:15215323"

/translations="ARGIIALLSLASIIIVVLLIKVLYFLCQPLHIFPRKQL

CDGLPCPLGEDEHCVKSPPEGFAVAVRLSKDRSTLQVLDSTGNWFGACPDNFEA

LAETACRMGYSSKPTFRAVEIGPDODLDVETENSQELRMNSSPCLSGSLVSLH

CLACGSLKTPRVVGGEEASVDSPQVSIQDKOHVCGSLDHPHVLTAHCFRKH

TDVNNKVRAGSKLGSFSLAVAKLIIIFENPMYFDNDIALMKLQFFTFSGTWRP

ICLPPDEELTPATPILGWFTKONGKMSDILLOASVQVIDSTRCDNADAYQGEV

TEKMCAGTPEGGVDTCQDSSGGLMYQSDQWVGVIVSGYCGGPFPGVTVKVA

YLNWIYNVWKAEL"

436 a 501 c 486 g 370 t

AGE COUNT

RIGIN

Alignment Scores:

Pred. No.: 8,49e-198 Length: 1793
 Score: 2167.00 Matches: 402
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 92.53% Indels: 0
 DB: 9 Gaps: 0

US-09-607-745-2 (1-435) x BC012752 (1-1793)

QY 34 ILeIleIleAlaLeuLeuSerLeuAlaSerIleIleIleValValValLeuLeuLysVal 53
 DB 11 ATCATCATAGCACTACTGAGCTGCGGAGTATCATTTGTGTGCTCATCAAGGTTG 70
 QY 54 ILeLeuAspLysTyrTrpPheLeuCysGlyGlnProLeuHisPheIleProArgLysGln 73
 DB 71 ATTCTGGATAAATACTACTTCTCTCGGGGAGCCTCTCCATCTCATCCGAGGAAGCAG 130
 QY 74 LeuCysAspGlyGluLeuAspCysProLeuGlyGluAspGluGluHisCysValLysSer 93
 DB 131 CTGTGTGACGGAGAGCTGGACTGTCCCTTTGGGGGAGGACGAGGAGCACTGTGTCAAGAGC 190
 QY 94 PheProGluGlyProAlaValAlaValArgLeuSerLysAspArgSerThrLeuGlnVal 113
 DB 191 TTCCCCGAAGGGCTTGCAGTGGAGTCCGCTCTCCAGGACCGATCCACACTGCGAGGTTG 250
 QY 114 LeuAspSerAlaThrGlyAsnTrpPheSerAlaCysPheAspAsnPheThrGluAlaLeu 133
 DB 251 CTGGACTCGCCACAGGGAACCTGTCTCTGCTGTCTTTCGACAACTTCACAGAAGCTCTC 310
 QY 134 AlaGluThrAlaCysArgGlnMetGlyTyrSerSerLysProThrPheArgAlaValGlu 153
 DB 311 GCTGAGACAGCCTGTAGGAGATGGGCTACAGCAGCAACCCCACTTTCAGAGCTGTGGAG 370
 QY 154 ILeGlyProAspGlnAspLeuAspValValGluIleThrGluAsnSerGlnGluLeuArg 173
 DB 371 ATGGCCCCAGACCAAGATCTGGATGTGTGTGAAATCACAGAAACACAGCAGGAGCTTCGC 430
 QY 174 MetArgAsnSerSerGlyProCysLeuSerGlySerLeuValSerLeuHisCysLeuAla 193
 DB 431 ATGGGAACCTCAAGTGGGCGCTCTCTCAGGCTCCCTGCTCTCCCTGCTCTCTTCTGCC 490
 QY 194 CysGlyLysSerLeuLysThrProArgValValGlyGlyGluGluLysValAspSer 213
 DB 491 TGTGGGAAGAGCCTCAAGACCCCGCTGTGGTGGGTGGGAGGAGCCCTCTGTGGATTCT 550
 QY 214 TrpProTrpGlnValSerIleGlnTyrAspLysGlnHisValCysGlyGlySerIleLeu 233
 DB 551 TGGCCTTGGCAGGTTCAGCATCCAGTACGACAAACAGCAGCATCTGTGTGGAGGAGCATCTCG 610
 QY 234 AspProHisTrpValLeuThrAlaAlaHisCysPheArgLysHisThrAspValPheAsn 253
 DB 611 GACCCCACTGGGTCTCTCAGCGCAGCCCACTGCTTCCAGGAAACATACCATGTGTTCAC 670
 QY 254 TrpLysValArgAlaGlySerAspLysLeuGlySerPheProSerLeuAlaValLys 273
 DB 671 TGGAGGTTCGGGCGAGGCTCAGCAAACTGGGCGCTTCCCACTCCCTGTGTGGCCAAAG 730
 QY 274 ILeIleIleIleGluPheAsnProMetTyrProLysAspAsnAspIleAlaLeuMetLys 293
 DB 731 ATCATCATATTGAAATTCACCCCACTGATACCCCAAGACATGACATCCCTCATGAAG 790
 QY 294 LeuGlnPheProLeuThrPheSerGlyThrValArgProIleCysLeuProPhePheAsp 313
 DB 791 CTGCAGTTCCTCACTTCTTTCAGGACAGTCAGGCCCATCTGTCTGCTGCTCTTCTTGTAT 850
 QY 314 GluGluLeuThrProAlaThrProLeuTrpIleIleGlyTrpGlyPheThrLysGlnAsn 333
 DB 851 GAGGAGCTCACTCCAGCCACCCCACTCTGATCATTTGATGGGGCTTTTACGAAGCAGAAAT 910
 QY 334 GlyGlyLysMetSerAspIleLeuGlnAlaSerValGlnValIleAspSerThrArg 353
 DB 911 GGAGGGAAGATGTCTGACATCTGCTGCGAGGCGTCAGTCCAGGTCAATTGACAGCACACGG 970

```
Y 354 CysAsnAlaAspAlaTyrGlnGlyGluValThrGluLysMetMetCysAlaGlyIle 373
b 971 TGC AATGCAGACGATGCGTAC CAGGGGGAAGTCACCGAGAAGATGATGTGCAGGCATC 1030
Y 374 ProGluGlyGlyValAspThrCysGlnGlyAspSerGlyGlyProLeuMetTyrGlnSer 393
b 1031 CCGAAGGGGGTGTGGACACCTGCCAGGGTGACAGTGTGGGGCCCTGATGTACCAATCT 1090
Y 394 AspGlnTrpHisValValGlyIleValSerTrpGlyTyrGlyCysGlyGlyProSerThr 413
b 1091 GACCAGTGGCATGTGTGGGCATCGTTAGCTGGGGCTATGGCTCGGGGGCCCGAGCACC 1150
Y 414 ProGlyValTyrThrLysValSerAlaTyrLeuAsnTrpIleTyrAsnValTrpLysAla 433
b 1151 CCAGGAGTATACACCAAGGTCTCAGCCTATCTCAACTGGATCTACAATGTCTGGAAGGCT 1210
Y 434 GluLeu 435
b 1211 GAGCTG 1216
```

Search completed: November 23, 2003, 10:39:57
Ob time : 5540.45 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

4 protein - nucleic search, using frame_plus_p2n model

run on: November 23, 2003, 07:53:55 ; Search time 87.359 Seconds
(without alignments)
2197.846 Million cell updates/sec

title: US-09-607-745-2

perfect score: 2342
sequence: 1 MPPDSQPLNSLDVPLRKP.....VYTKSAYLNWYVWKAEL 435

scoring table:

BLOSUM62

Xgapop 10.0 , Xgapext 0.5

Ygapop 10.0 , Ygapext 0.5

Fgapop 6.0 , Fgapext 7.0

Delop 6.0 , Delext 7.0

searched: 569978 seqs, 220691566 residues

oral number of hits satisfying chosen parameters: 1139956

inimum DB seq length: 0

aximum DB seq length: 2000000000

ost-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

command line parameters:

MODEL=frame+ p2n.model -DBV=xlh
O=/cgn2_1/USPTO.spool/US09607745/runat 21112003 144346 22244/app query.fasta_1.1038
Q=Issued Patents NA -QFMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0
LOOPEXT=0 -UNITS=BITS -START=1 -END=1 -MATRIX=blomsum62 -TRANS=human40.cdi
LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
USER=US09607745 @CGN 1.1 76 @runat 21112003 144346 22244 -NCPU=6 -ICPU=3
MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DGPBLOCK=100 -LONGLOG
DEVT TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

database :

Issued Patents NA.*
1: /cgn2_6/ptodata/2/ina/5A COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PTCUS COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

result No.	Score	Query Match	Length	ID	Description
1	2338	99.8	2038	3	US-09-008-271A-18
2	2324	99.2	2079	4	US-09-656-002-1
3	677.5	28.9	2479	3	US-09-342-749-29
4	677.5	28.9	2479	4	US-09-691-840-29
5	676.5	28.9	1479	3	US-09-342-749-1
6	676.5	28.9	1479	4	US-09-691-840-1
7	650.5	27.8	2413	3	US-09-518-046-1
8	627	26.8	2544	3	US-09-518-046-3
9	621.5	26.5	2416	3	US-09-261-416-1
10	584.5	25.0	1783	3	US-09-510-738A-188
11	584.5	25.0	1783	4	US-09-861-966-188
12	584.5	25.0	2363	4	US-09-742-703-3

13	580	24.8	1615	4	US-09-820-002-1	Sequence 1, Appl
14	578	24.7	1605	2	US-09-000-846-1	Sequence 1, Appl
15	576	24.6	1460	4	US-09-370-838-80	Sequence 10, Appl
16	576	24.6	1517	1	US-08-508-448C-15	Sequence 85, Appl
17	576	24.6	2790	4	US-09-370-838-79	Sequence 79, Appl
18	574	24.5	1462	4	US-09-370-838-55	Sequence 55, Appl
19	571	24.4	2581	1	US-08-200-900A-1	Sequence 1, Appl
20	571	24.4	2581	5	PCT-US94-00616-1	Sequence 1, Appl
21	558.5	23.8	1077	3	US-08-807-151-2	Sequence 2, Appl
22	558.5	23.8	1077	4	US-09-478-957-2	Sequence 2, Appl
23	531.5	22.7	696	1	US-08-508-448C-24	Sequence 24, Appl
24	516.5	22.1	901	1	US-08-508-448C-9	Sequence 9, Appl
25	503.5	21.5	959	4	US-09-023-942A-25	Sequence 25, Appl
26	496	21.2	1613	4	US-09-387-375-1	Sequence 1, Appl
27	483	20.6	1100	4	US-09-023-942A-5	Sequence 5, Appl
28	478	20.4	1081	3	US-09-008-271A-15	Sequence 15, Appl
29	477	20.4	1094	4	US-09-023-942A-3	Sequence 3, Appl
30	472.5	20.2	1430	4	US-09-386-629-1	Sequence 1, Appl
31	467.5	20.0	1110	4	US-09-386-653A-1	Sequence 1, Appl
32	466.5	19.9	1225	4	US-09-734-675-1	Sequence 1, Appl
33	463	19.8	3147	2	US-09-027-337-1	Sequence 1, Appl
34	463	19.8	3147	4	US-09-644-600-1	Sequence 1, Appl
35	463	19.8	3147	4	US-09-644-600-18	Sequence 18, Appl
36	462.5	19.7	980	4	US-09-023-942A-30	Sequence 30, Appl
37	462.5	19.7	1212	4	US-09-620-312D-431	Sequence 431, App
38	458.5	19.6	1130	4	US-09-387-375-8	Sequence 8, Appl
39	458	19.6	1739	2	US-08-681-151-2	Sequence 2, Appl
40	449.5	19.2	1165	4	US-09-023-942A-28	Sequence 28, Appl
41	445	19.0	933	4	US-09-023-942A-29	Sequence 29, Appl
42	443.5	18.9	1130	4	US-09-386-653A-8	Sequence 8, Appl
43	443	18.9	2296	1	US-07-750-080A-18	Sequence 18, Appl
44	443	18.9	2296	3	US-08-651-472-18	Sequence 18, Appl
45	443	18.9	2296	3	US-08-358-928-18	Sequence 18, Appl

ALIGNMENTS

RESULT 1
US-09-008-271A-18
; Sequence 18, Application US/09008271A
; Patent No. 6203979
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; Hillman, Jennifer L.
; Yue, Henry
; Guegler, Karl J.
; Corley, Neil C.
; Tang, Tom Y.
; Shah, Purvi
; TITLE OF INVENTION: HUMAN PROTEASE MOLECULES
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Dr.
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/008,271A
; FILING DATE: 16-Jan-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: <Unknown>
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Mohan-Peterson, Sheela
; REGISTRATION NUMBER: 41,201
; REFERENCE/DOCKET NUMBER: PP-0458 US


```
B: 4 Gaps: 0
S-09-607-745-2 (1-435) x US-09-656-002-1 (1-2079)
2 AspProAspSerAspGlnProLeuAsnSerLeuAspVallysProLeuArgLysProArg 21
217 GATCCTGACAGTATCAACCTCTGAACAGAGCTCGATGTCAACCCCTGCCAACCCTG 276
22 IleProMetGluThrPheArgLys-ValGlyIleProIleIleIleAlaLeuSerLe 41
277 ATCCCATGGAGACCTTCAGAAAGTGTGGGATCCCATCATCATAGCACTACTGAGCCT 336
41 uAlaSerIleIleIleValValValLeuIleLysValIleLeuAspLysTyrPheLe 61
337 GCGAGTATCATATGTTGTTCCTTCACAGGTGATTCGGATAAATACTACTTCCT 396
61 uCysGlyGlnProLeuHisPheLeuProArgLysGlnLeuCysAspGlyGluLeuAspCy 81
397 CTGGGGGACGCTCTCCACTTCATCCGAGAGACAGCTGTGTGACGAGAGCTGGACTG 456
81 sProLeuGlyGluAspGluGluHisCysValLysSerPheProGluGlyProAlaValAl 101
457 TCCTTGGGGGAGACAGAGAGACTGTGTCAAGAGCTTCCCGAAGGGCTCGAGTGGC 516
101 aValArgLeuSerLysAspArgSerThrLeuGlnValLeuAspSerAlaThrGlyAsnTr 121
517 AGTCGGCTCTCCAAAGACCGATCCACACTGCAGGTGCTGACTCGGCCACAGGAACTG 576
121 pPheSerAlaCysPheAspAsnPheThrGluAlaLeuAlaGluThrAlaCysArgGlnMe 141
577 GTTCTCTGCTGTTTTCACAACTTCACAGAACTCTCGCTGACAGACGCTGTAGGCAGAT 636
141 tGlyTyrSerSerLysProThrPheArgAlaValGluIleGlyProAspGlnAspLeuAs 161
637 GGCTACAGCAGCAAAACCCACTTTCAGAGCTGTGGAGATTGGCCCGACAGCACTCGGA 696
161 pValValGluIleThrGluAsnSerGlnGluLeuArgMetArgAsnSerSerGlyProCy 181
697 TGTGTTGTAATCACAGAAACAGCAGGAGCTTCGCATCGGAACTCAAGTGGGCCCTG 756
181 sLeuSerGlySerLeuValSerLeuHisCysLeuAlaCysGlyLysSerLeuLysThrPr 201
757 TCTCTCAGGCTCCTGCTCCTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 816
201 oArgValValGlyGlyGluAlaSerValAspSerTyrProTyrGlnValSerIleG 221
817 CCGTGTGTTGGTGGGGAGAGAGCCTCTGTGGATTCCTTGGCCTTGGCAGGTCAGCATCCA 876
221 nTyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThrAl 241
877 GTACGCACAAACAGCAGCTCTGTGGAGGAGCATCTGGACCCCATCTGGGTCTCTCAGGC 936
241 aAlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySerAs 261
937 AGCCCACTGCTTCAGGAAACATACCGATGTGTTCACCTGGAAGTGGCGGACGGCTCAGA 996
261 pLysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleIleGluPheAsnPr 281
997 CAACTGGGCGAGCTTCCCATCCCTGGCTGTGGCCAAAGATCATCATTTGAATTCACCC 1056
281 oMetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPheSe 301
1057 CATGTACCCCAAGACATATGACATCGCCCTCATGAGCTGCATGTTCCCATCTCACTTCTC 1116
301 rGlyThrValArgProIleCysLeuProPhePheAspGluGluLeuThrProAlaThrPr 321
1117 AGGCACAGTCAGGCCATCTGTCTGCCCTCTTGTGATGAGAGACTCACTCCAGCCACCCC 1176
321 oLeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyGlyLysMetSerAspIleLe 341
1177 ACTCTGGATCATTTGGATGGGCTTTTCAGAACGAGAAATGGAGGAGAGATGTCTGACATACT 1236
341 uLeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAlaTyrG 361
```

Db 1237 GCTCCAGGCGTCAGTCCAGTCAATTCACAGCACACGGTCAATTCGACACCATCGTACCA 1296

QY 361 nGlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThrCy 381

Db 1297 GGGGGAAGTCACCGAGAGATGATGTGTGCAGGATCCCGGAAGGGGTGTGGACACCTG 1356

QY 381 sGlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGlyI 401

Db 1357 CCAGGCTGACAGTGGTGGGCCCTGATGTACCAATCTGACCACTGGCATGTGCTGGGCAT 1416

QY 401 eValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyValTyrThrLysValSe 421

Db 1417 CGTTAGCTGGGCTATGGCTGCGGGGGCCCCAGCACCCAGGAGTATACACCAAGGTC 1476

QY 421 rAlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu 435

Db 1477 AGCCTATCTCACTGGATCTTACAAATGCTTGGAGGCTGAGCTG 1519

RESULT 3

US-09-342-749-29

; Sequence 29, Application US/09342749

; Patent No. 6166194

; GENERAL INFORMATION:

; APPLICANT: Wong, Alexander K.C.

; APPLICANT: Tavtigian, Sean V.

; APPLICANT: Teng, David H.-F.

; APPLICANT: Myriad Genetics, Inc.

; TITLE OF INVENTION: TMRSS2 is a Tumor Suppressor

; FILE REFERENCE: 2318-202

; CURRENT APPLICATION NUMBER: US/09/342,749

; CURRENT FILING DATE: 1999-06-29

; EARLIER APPLICATION NUMBER: US 60/091,044

; EARLIER FILING DATE: 1998-06-29

; NUMBER OF SEQ ID NOS: 33

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 29

; LENGTH: 2479

; TYPE: DNA

; ORGANISM: Homo sapiens

US-09-342-749-29

Alignment Scores:

Pred. No.: 1,816-63 Length: 2479

Score: 677.50 Matches: 151

Percent Similarity: 54.05% Conservative: 56

Best Local Similarity: 39.43% Mismatches: 127

Query Match: 28.93% Indels: 49

DB: 3 Gaps: 13

US-09-607-745-2 (1-435) x US-09-342-749-29 (1-2479)

QY 75 CysAspGlyGluLeuAspCysProLeuGlyGluAspGluHisCysValLysSerPhe 94

Db 453 TGTGATGGGTCTCACACTGCCCGGGGGAGGACAGAAATCGGTGTGTTCCTCTTAC 512

QY 95 ProGluGlyProAlaValAlaValArgLeuSerLysAspArgSerThrLeuGlnValLeu 114

Db 513 -----GGACCA-----AACTTCATCCTCTCAGATGTAC 539

QY 115 AspSerAlaThrGlyAsnTrpPheSerAlaCysPheAspAsnPheThrGluAlaLeuAla 134

Db 540 TCATCTCAGAGAAAGTCTCGGCCCTGTGTGCCAAGACGACTGGAAACGAGAACTACCGG 599

QY 135 GluThrAlaCysArgGlnMetGlyTyrSerSerLysProThrPheArgAlaValGluIle 154

Db 600 CCGGGCGCTCTCAGGAGACATGGGCTAT-----AGAATAATTTTACTCTAGCCAA--- 650

QY 155 GlyProAspGlnAspLeuAspValValGluIleThrGluAsnSerGlnGluLeuArgMet 174

Db 651 -----GGAATAGTGGATGACAGCGGATCCACAGCTTTATGAACTG 692

QY 175 ArgAsnSerSerGly-----ProCysLeu 182

[illegible]

D _b	933	ACAGCGCCACATGCGTGGAAAAACCTCTTAA	CAATCCATGCATTTGGACGGCATTTGGCG	992
Q _y	259	GlySerAspLysLeuGlySerPhe	-----ProSerLeuAlaValAlaLysIle	274
D _b	993	GGGATTTTGAGA--CAATCTTTTCATGTTCT	TATGGAGCCGGATACCAAGTACAAAAAGTG	1049
Q _y	275	IleIleIleGluPheAsnProMetTyr	-----ProLysAspAsnAspIleAlaLeu	291
D _b	1050	ATT-----TCTCATCAAAATTTGACT	CCAAGACCAAGAACCAATGACATTCGCGCTG	1100
Q _y	292	MetLysLeuGlnPheProLeuThrPheSerGlyThr	ValArgProIleCysLeuProPhe	311
D _b	1101	ATCAAGCTGAGAAGCGCTCTGACTTTTCAACG	CACTAGTGAAACCAAGTGTGTCTGCCCAAC	1160

Y 312 PheAspGluGluLeuThrProAlaThrProLeuTrpIleileGlyTrpGlyPheThrLys 331
b 1161 CGAGGATGATGCTGCAGCAGAACAGCTCTGCTGGATTTCGGGGTGGGGGCCACCGAG 1220
Y 332 GlnAsnGlyGlyLysMetSerAspIleLeuLeuGlnAlaSerValGlnValIleAspSer 351
b 1221 GAGAAA--GGGAGACCTCAGAGTCTGAACGCTGCCAAGGTCTCTCATTTGAGACA 1277
Y 352 ThrArgCysAsnAlaAspAspAlaTyrGlnGlyValThrGluLysMetMetCysAla 371
b 1278 CAGAGATGCAACGACAGATGCTCTATGACAACCTGATCATCACACGACCATGATCTGTGCC 1337
Y 372 GlyLeuProGluGlyValAlaSerThrCysGlnGlyAspSerGlyGlyProLeuMetTyr 391
b 1338 GGTCTTCTGCAGGGGAACGTCGATTCTTGCAGGGTGACAGTGGAGGGCCTCTGGTCACT 1397
Y 392 GlnSerAspGln---TrpHisValValGlyIleValSerTrpGlyTyrGlyCysGlyGly 410
b 1398 TCGAACACAAATATCTGGTGGCTGATAGGGGATACAAAGCTGGGGTCTGGCTGTGCCAAA 1457
Y 411 ProSerThrProGlyValTyrThrLysValSerAlaTyrLeuAsnTrpIleTyrAsnVal 430
b 1458 GCTTACAGACCAAGAGTGTACGGGAATGTGATGATATTCACGAGCTGGATTATCGACAA 1517
Y 431 TrpLysAla 433
b 1518 ATGAGGCA 1526

RESULT 5
S-09-342-749-1
Sequence 1, Application US/09342749
Patent No. 6166194
GENERAL INFORMATION:
APPLICANT: Wong, Alexander K.C.
APPLICANT: Tavtigian, Sean V.
APPLICANT: Teng, David H.-F.
APPLICANT: Myriad Genetics, Inc.
TITLE OF INVENTION: TMPSR2 is a Tumor Suppressor
FILE REFERENCE: 2318-202
CURRENT APPLICATION NUMBER: US/09/342,749
CURRENT FILING DATE: 1999-06-29
EARLIER APPLICATION NUMBER: US 60/091,044
EARLIER FILING DATE: 1998-06-29
NUMBER OF SEQ ID NOS: 33
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1
LENGTH: 1479
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (1)..(1476)
FEATURE:
NAME/KEY: conflict
LOCATION: (724)
OTHER INFORMATION: Listed as T in GenBank Accession No. U75329
FEATURE:
NAME/KEY: conflict
LOCATION: (985)
OTHER INFORMATION: Listed as C in GenBank Accession No. 6166194 U75329
FEATURE:
NAME/KEY: conflict
LOCATION: (1347)
OTHER INFORMATION: Listed as C in GenBank Accession No. 6166194 U75329
FEATURE:
NAME/KEY: conflict
LOCATION: (1466)
OTHER INFORMATION: Listed as A in GenBank Accession No. 6166194 U75329
FEATURE:
NAME/KEY: conflict
LOCATION: (1471)
OTHER INFORMATION: Listed as A in GenBank Accession No. 6166194 U75329.
FEATURE:

; NAME/KEY: allele
; LOCATION: (478)
; OTHER INFORMATION: This base can be G or A with G being the more
; OTHER INFORMATION: common allele. The codon will change from Val to
; OTHER INFORMATION: Met.
; FEATURE:
; NAME/KEY: allele
; LOCATION: (777)
; OTHER INFORMATION: This base can be C or T with C being the more
; OTHER INFORMATION: common allele. The codon is unaffected with both
; OTHER INFORMATION: alleles encoding Gly.
; FEATURE:
; NAME/KEY: allele
; LOCATION: (768)
; OTHER INFORMATION: This base can be C or T with C being the more
; OTHER INFORMATION: common allele. This is a silent polymorphism.
; FEATURE:
; NAME/KEY: allele
; LOCATION: (834)
; OTHER INFORMATION: This base can be C or T with C being the more
; OTHER INFORMATION: common allele. This is a silent polymorphism.
; FEATURE:
; NAME/KEY: allele
; LOCATION: (625)
; OTHER INFORMATION: This base can be T or A with T being the more
; OTHER INFORMATION: common allele. The codon will change from Phe to
; OTHER INFORMATION: Ile
US-09-342-749-1

Alignment Scores:

Pred. No.:	1,04e-63	Length:	1479
Score:	676.50	Matches:	150
Percent Similarity:	53.91%	Conservative:	57
Best Local Similarity:	39.06%	Mismatches:	128
Query Match:	28.89%	Indels:	49
DB:	3	Gaps:	13

US-09-607-745-2 (1-435) x US-09-342-749-1 (1-1479)

QY	75	CysAspGlyGluLeuAspCysProLeuGlyGluAspGluGluHisCysValLysSerPhe	94
Db	397	TGTGATGGGTGTACACTGCCCGCGGGAGGACGAGAAATCGGTGTGTTCGCTCTAC	456
QY	95	ProGluGlyProAlaValAlaValArgLeuSerLysAspArgSerThrLeuGlnValLeu	114
Db	457	-----GGACCA-----AACTTCATCCTCTCAGGTGTAC	483
QY	115	AspSerAlaThrGlyAsnTrpPheSerAlaCysPheAspAsnPheThrGluAlaLeuAla	134
Db	484	TCATCTCAGAGGAAGTCTGGCACCTGTGTGCCAAGACGACTGGAACTGAGAACTACGGG	543
QY	135	GluThrAlaCysArgGlnMetGlyTyrSerSerLysProThrPheArgAlaValGluIle	154
Db	544	CGGGCGGCTCGCAGGACATGGCTAT-----AAGAATAATTTTACTCTAGCCAA---	594
QY	155	GlyProAspGlnAspLeuAspValValGluIleThrGluAsnSerGlnGluLeuArgMet	174
Db	595	-----GGAATAGTGGATGATGACGCGGATCCACCAGCTTTATGAACATG	636
QY	175	ArgAsnSerSerGly-----ProCysLeu	182
Db	637	AACACAAAGTCCGGCAATGCTGATATCTATAAAACTGATACACAGTATGCTCTTCT	696
QY	183	SerGlySerLeuValSerLeuHisCysLeuAlaCysGlyLysSerLeu-----Lys	199
Db	697	TCAAAGCAGTGGTTCCTTAGCTGTATAGCTGCGGGTCACTTGAACCAACCGCG	756
QY	200	ThrProArgValValGlyGlyGluAlaSerValAspSerTrpProTrpGlnValSer	219
Db	757	CAGAGCAGGATGCTGGCGCGAGAGCGGCTCCCGGGGGGCTGGCCCTTGGCAGGTGAGC	816
QY	220	IleGlnTyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeu	239

817 CTGACGCTCCAGAACTCCACGCTGCGGAGGCTCCATCATCACCCCGAGTGGTCTG 876
240 ThrAlaAlaHisCysPheArgLysHis---ThrAspValPheAsnTrpLysValArgAla 258
877 ACAGCGCGCCACTGCGTGAAAGAACTCTTAACAATCCATGCGATGGAGCGCATTTGG 936
259 GlySerAspLysLeuGlySerPhe-----ProSerLeuAlaValAlaLysIle 274
937 GGGATTTCAGAA---CAATCTTTCATGTTCTTANGAGCGGATACCAAGTGAAGAAAGTG 993
275 IleIleIleGluPheAsnProMetTyr-----ProLysAspAsnAspIleAlaLeu 291
994 ATT-----TCTCATCCAAATATGATCTCAAGACCAAGACCAATGACATTCGGTG 1044
292 MetLysLeuGlnPheProLeuThrPheSerGlyThrValArgProIleCysLeuProPhe 311
1045 ATGAAGCTGCAGAGCCCTGACTTCAACGACCTAGTGAAACCACTGTGTCTGCCAAC 1104
312 PheAspGluGluLeuThrProAlaThrProLeuTrpIleIleGlyTrpGlyPheThrLys 331
1105 CCAGCATGATGCTGCAGCCAGAAACAGCTCTCTGCTGATTTCCGGGTGGGGCCACCGAG 1164
332 GlnAsnGlyGlyLysMetSerAspIleLeuLeuGlnAlaSerValGlnValIleAspSer 351
1165 GAGAAA---GGGAACACCTCAGAACTGTGAACGCTGCCAAGTGTCTTCTATTGAGACA 1221
352 ThrArgCysAsnAlaAspAlaTyrGlnGlyValThrGlnGlyValThrGlyLysMetCysAla 371
1222 CAGAGATCAACAGCAGATATGCTATGACAACTGATCACCAGCCATGATCTGTGCC 1281
372 GlyIleProGluGlyGlyValAspThrCysGlnGlyAspSerGlyGlyProLeuMetTyr 391
1282 GCTTCCTCGCGGGAAGCTGATTTGCCAGGGTGACAGTGGAGGGCCCTCTGTCTACT 1341
392 GlnSerAspGln---TrpHisValValGlyIleValSerTrpGlyTyrGlyCysGlyGly 410
1342 TCRAAGAACATATCTGTTGGCTGATAGGGGATACAGCTGGGTCTTGGCTGTGCCAAA 1401
411 ProSerThrProGlyValTyrThrLysValSerAlaTyrLeuAsnTrpIleTyrAsnVal 430
1402 GCTTACAGACACAGGAGTGTACGGGAATGTGATGTATTACAGGACTGGATTATCGCAA 1461
431 TrpLysAlaGlu 434
1462 ATGAGGGCGAGAC 1473

RESULT 6

US-09-691-840-1

Sequence 1, Application US/09691840

Patent No. 644419

GENERAL INFORMATION:

APPLICANT: Wong, Alexander K.C.

APPLICANT: Tavtigian, Sean V.

APPLICANT: Teng, David H.-F.

APPLICANT: Myriad Genetics, Inc.

TITLE OF INVENTION: TWFRS2 is a Tumor Suppressor

FILE REFERENCE: 2318-202

CURRENT APPLICATION NUMBER: US/09/691,840

PENDING FILING DATE: 2000-10-18

PRIOR APPLICATION NUMBER: US/09/342,749

PRIOR FILING DATE: 1999-06-29

PRIOR APPLICATION NUMBER: US 60/091,044

PRIOR FILING DATE: 1998-06-29

NUMBER OF SEQ ID NOS: 33

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 1

LENGTH: 1479

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: CDS

LOCATION: (1)..(1476)

NAME/KEY: conflict

LOCATION: (724)
OTHER INFORMATION: Listed as T in GenBank Accession NO. U75329
NAME/KEY: conflict
LOCATION: (985)
OTHER INFORMATION: Listed as C in GenBank Accession No. 6444419 U75329
NAME/KEY: conflict
LOCATION: (1347)
OTHER INFORMATION: Listed as C in GenBank Accession No. 6444419 U75329
NAME/KEY: conflict
LOCATION: (1466)
OTHER INFORMATION: Listed as A in GenBank Accession No. 6444419 U75329
NAME/KEY: conflict
LOCATION: (1471)
OTHER INFORMATION: Listed as A in GenBank Accession No. 6444419 U75329.
NAME/KEY: allele
LOCATION: (478)
OTHER INFORMATION: This base can be G or A with G being the more common allele. The codon will change from Val to Met.
OTHER INFORMATION: Met.
NAME/KEY: allele
LOCATION: (777)
OTHER INFORMATION: This base can be C or T with C being the more common allele. The codon is unaffected with both alleles encoding Gly.
OTHER INFORMATION: Met.
NAME/KEY: allele
LOCATION: (768)
OTHER INFORMATION: This base can be C or T with C being the more common allele. This is a silent polymorphism.
NAME/KEY: allele
LOCATION: (834)
OTHER INFORMATION: This base can be C or T with C being the more common allele. This is a silent polymorphism.
NAME/KEY: allele
LOCATION: (625)
OTHER INFORMATION: This base can be T or A with T being the more common allele. The codon will change from Phe to Ile
US-09-691-840-1

Alignment Scores:

Pred. No.:	1,04e-63	Length:	1479
Score:	676.50	Matches:	150
Percent Similarity:	53.91%	Conservative:	57
Best Local Similarity:	39.06%	Mismatches:	128
Query Match:	28.89%	Indels:	49
DB:	4	Gaps:	13

US-09-607-745-2 (1-435) x US-09-691-840-1 (1-1479)

Qy	75	CysAspGlyGluLeuAspCysProLeuGlyGluAspGluHisCysValLysSerPhe	94
Db	397	TGTGATGGCGTGCACACTGCCCGGGGAGCGAGAAATCGGTGTTCGCTCTAC	456
Qy	95	ProGluGlyProAlaValAlaValArgLeuSerLysAspArgSerThrLeuGlnValLeu	114
Db	457	-----GGACCA-----	483
Qy	115	AspSerAlaThrGlyAsnTrpPheSerAlaCysPheAspAsnPheThrGluAlaLeuAla	134
Db	484	TCATCTCAGAGGAAGTCTGGCACCTGTGTGCCAAGCAGCTGGACGAACTACGGG	543
Qy	135	GluThrAlaCysArgGlnMetGlyTyrSerLysProThrPheArgAlaValGluIle	154
Db	544	CGGCGGCGCTGCAGGACATGGGCTAT-----AAGAATAATTTTACTCTAGCCAA--	594
Qy	155	GlyProAspGlnAspLeuAspValGluIleThrCluAsnSerGlnGluLeuArgMet	174
Db	595	-----GGATATAGTGGATGACAGCGGATCCACCGAGTTTATGAAACTG	636
Qy	175	ArgAsnSerSerGly-----	182
Db	637	AAACAAGTCCCGGCATGTCGATATCTATAAAAACTGTACACAGTGTGCTGTCT	696
Qy	183	SerGlySerLeuValSerLeuHisCysLeuAlaCysGlyLysSerLeu-----Lys	199

```
697 TCAAAGCAGTGGTTCTTACCGTGTATAGCTCGCGGTCACTGAACCTCAAGCCGC 756
200 ThrProArgValValGlyGlyGluAlaSerValAspSerTrpProTrpGlnValSer 219
757 CAGAGCAGGATCGTGGCGCGGAGAGCGGCTCCCGGGGCGCTGGCCCTGGCAGGTG 816
220 IleGlnTrpAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeu 239
817 CTGCAGCTCCAGAACGCTCCAGCTGTGGGAGGTCTCCATCATCCCGGAGTGGATCGT 876
240 ThrAlaAlaHisCysPheArgLysHis---ThrAspValPheAsnTrpLysValArgAla 258
877 ACGCCGCCCACTCGCTGGTGAACACCTTTAAACATCCATGGCATTTGGACGGCATTTGG 936
259 GlySerAspLysLeuGlySerPhe-----ProSerLeuAlaValAlaLysIle 274
937 GGGATTTTGAGA--CAATCTTTCATGTTCTATGGAGCGCGATACCAAGTAGAAGTAG 993
275 IleIleIleGluPheAsnProMetTyr-----ProLysAspAsnAspIleAlaLeu 291
994 ATT-----TCTCATCCAAATTATGACTCCAAGACCAAGAACCAATGACATTGGCGTG 1044
292 MetLysLeuGlnPheProLeuThrPheSerGlyThrValArgProIleCysLeuProPhe 311
1045 ATGAAGCTGCAGAGCTCTGACTTTCAAGACCTAGTGAACACAGTGTCTGCCCCAAC 1104
312 PheAspGluGluLeuThrProAlaThrProLeuTrpIleIleGlyTrpGlyPheThrLys 331
1105 CCAGGCATGATGTCGAGCAGACAGCTCTGCTGGATTTCGCGGTGGGGGCCACCGAG 1164
332 GlnAsnGlyGlyLysMetSerAspIleLeuLeuGlnAlaSerValGlnValIleAspSer 351
1165 GAGAAA---GGGAGAACCTCAGAAAGTGTGAACGCTGCCAAGGTGCTTCTCATTTGAGACA 1221
352 ThrArgCysAsnAlaAspAlaTyrGlnGlyGluValThrGluLysMetMetCysAla 371
1222 CAGAGATGCAACAGCAGATATGCTATGACAACTGATCACACACCCCATGATCTGGC 1281
372 GlyIleProGluGlyValAspThrCysGlnGlyAspSerGlyGlyProLeuMetTyr 391
1282 GGTCTCTCGAGGGGACGTCGATTCTTGCCAGGTGACAGTGGAGGGCTCTGGTCACT 1341
392 GlnSerAspGln---TrpHisValValGlyValSerTrpGlyTyrGlyCysGlyGly 410
1342 TCGAAGAACAAATATCTGGTGGCTGTAGAGGATACAAAGCTGGGGTTCTGCTGTGCCAAA 1401
411 ProSerThrProGlyValThrLysValSerAlaTyrLeuAsnTrpIleTyrAsnVal 430
1402 GCTTACAGACAGGAGTACGGGAATGTGATGTATTCAGGACTGGATTATCGACAA 1461
431 TrpLysAlaGlu 434
1462 ATGAGGGCAGAC 1473
```

RESULT 7

US-09-518-046-1

Sequence 1, Application US/09518046

Patent No. 6294663

GENERAL INFORMATION:

APPLICANT: O'Brien, Timothy J.

APPLICANT: Underwood, Lowell J.

TITLE OF INVENTION: Transmembrane Serine Protease Overexpressed

FILE OF INVENTION: in Ovarian Carcinoma and Uses Thereof

FILE REFERENCE: D6192CIP

CURRENT APPLICATION NUMBER: US/09/518,046

CURRENT FILING DATE: 2000-03-02

EARLIER APPLICATION NUMBER: 09/261,416

EARLIER FILING DATE: 1999-03-03

NUMBER OF SEQ ID NOS: 153

SEQ ID NO 1

LENGTH: 2413

TYPE: DNA

```
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; OTHER INFORMATION: entire cDNA sequence of TADG-12 gene
US-09-518-046-1
```

Alignment Scores:

```
Pred. No.: 1.49e-60 Length: 2413
Score: 650.50 Matches: 157
Percent Similarity: 52.14% Conservative: 74
Best Local Similarity: 35.44% Mismatches: 149
Query Match: 27.78% Indels: 63
DB: 3 Gaps: 18
```

US-09-607-745-2 (1-435) x US-09-518-046-1 (1-2413)

```
QY 22 IleProMetGluThrPheArgLysValGlyIleProIleIleAlaLeuSerLeu 41
DB 273 CTGCCATTTGAAGTTTTTCCCAATCATCGTCAATGGGGATCATTTGCATTTG----- 323
QY 42 AlaSerIleIleValValValLeuIleLysValIleLeuAsp-----LysTyr 58
DB 324 -----ATATTAGCACTGGCCATTGTCTGGGCATCCACTTCGACTGCTCAGGGAAGTAC 377
QY 59 -----TyrPheLeuCysGlyGlnProLeuHisPheIleProArgLysGlnLeu 74
DB 378 AGATGTCGCTCATCTCTTTAAGTGT-----ATCGAGCTGATACTCGA----- 419
QY 75 CysAspGlyGluLeuAspCysProLeuGlyGluAspGluGluHisCysValLysSerPhe 94
DB 420 TGTGACGAGTCTCGATTGCAAGCGGGAGGAGGAGTACCGCTGT----- 467
QY 95 ProGluGlyProAlaValAlaValArgLeuSerLysAspArgSerThrLeuGlnValLeu 114
DB 468 -----GTCCGGGTGGGTGTGTGAGAAATGCCGTGCTCCAGGTTTC 506
QY 115 AspSerAlaThrGlyAsnTrpPheSerAlaCysPheAspAsnPheThrGluAlaLeuAla 134
DB 507 ACAGTGTGT-----TCGTGGAAGACCATGTCTCCGATGACTGGAAGGTCTACTACGCA 560
QY 135 GluThrAlaCysArgGlnMetGlyTyrSerSer----- 145
DB 561 AATGTTGCTGTGTCACCACTGGGTTCCTCAAGTATGTGAGTTCAGATAACCTCAGAGTG 620
QY 146 -----LysProThrPheArgAlaValGluIleGlyProAspGlnAspLeuAspVal 162
DB 621 AGCTCGCTGGAGGGCAGTTCGGGAGGAGTGTGTCTCCATCGATCACCTCTGCCAGAT 680
QY 163 ValGluIleThrGluAsnSerGlnGluLeuArgMetArgAsnSerSerGlyProCysLeu 182
DB 681 GACAAGGTGACTGCTATTACCCACTCAGTATATGTGAGGGAGGGA-----TGTGCC 731
QY 183 SerGlySerLeuValSerLeuHisCysLeuAlaCysGly---LysSerLeuLysThrPro 201
DB 732 TCTGCCACGCGGTACTCTTGCAGTGCACAGCTGTGTCATAGAGGGGCTCAGCTCA 791
QY 202 ArgValValGlyGlyGluGluAlaSerValAspSerTrpProTrpGlnValSerIleGln 221
DB 792 CGCATCGTGGGTGGAACATGTCCTTGTCTCTCGAGTGGCCCTGGCAGGCCAGCCTTCAG 851
QY 222 TyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThrAla 241
DB 852 TTCCAGGGCTACCACTCTGTGGGGGCTCTGTCTCATCACGCCCTGTGGATCATCTGCT 911
QY 242 AlaHisCysPheArgLysHisThrAspValPhe-----AsnTrpLysValArgAla 258
DB 912 GCACACTTGTGT-----TATGACTTGTATCTCCCAAGTCAATGACCATCCAGGTG 962
QY 259 GlySerAspLysLeu-----GlySerPheProSerLeuAlaValAlaLysIleIle 276
DB 963 GGTCTAGTTTCCCTGTTGGACAATCCAGCCCATCCACTTGGTGGAGAGATTGTC--- 1019
QY 277 IleGluPheAsnProMetTyr---ProLys-----AspAsnAspIleAlaLeuMetLys 293
```

QY		42	AlaSerIlellelleValValValValLeulleLysVallelleuAsp-----LysTyr 58
Db		324	-----ATATTAGCATGCCACTGTGCTGGGCACCTTCGACTCGCTCAGGAAGTAC 377
QY		59	-----TyrPheLeuCysGlyGlnProLeuHisPheIleProArgLysGlnLeu 74
Db		378	AGATGTCGCTCACTCTTTAAGTG-----ATCAGCTGATAACTCGA----- 419
QY		75	CysAspGlyGluleuAspCysProLeuGlyGluAspGluGluHisCysValLysSerPhe 94
Db		420	TGTGACGGAGTCTCGGATTCCAAAGACGGGGAGCACGAGTACCCTGT----- 467
QY		95	ProGluGlyProAlaValAlaValArgLeuSerLysAspArgSerThrLeuGlnValLeu 114
Db		468	-----GFCGCGGTGGGTGCAGAANTCCGTGCTCCAGGTGTTTC 506
QY		115	AspSerAlaThrGlyAsnTrpPheSerAlaCysPheAspAsnPheThrGluAlaLeuAla 134
Db		507	ACAGCTGCT-----TCGTGGAGACCATGTGCTCCGATGACTGGAGGGTCACTACGCCA 560
QY		135	GluThrAlaCysArgGlnMetGlyTyrSerSer----- 145
Db		561	AATGTTGCCTGTGCCCACTGGGTTCCTCCAAAGCTATGTAACTTCAGATAACCTCAGAGTG 620
QY		146	-----LysProThrPheArgAlaValGluIleGlyProAspGlnAspLeuAspVal 162
Db		621	AGCTCGCTGGGGGGCAGTTTCGGGGAGGAGTTTGTGTCCATCGATCACTCTTCCACAGAT 680
QY		163	ValGluIleThrGluAsnSerGlnGluLeuArgMetArgAsnSerSerGlyProCysLeu 182
Db		681	GACAAGTGACTGCATTACCACTCAGTATATGTGAGGGAGGGA-----TGTCGCC 731
QY		183	SerGlySerIleValSerLeuHisCysLeuAlaCysGly---LysSerLeuLysThrPro 201

F-09-518-046-3
 Sequence 3, Application US/09518046
 Patent No. 6294663
 GENERAL INFORMATION:
 APPLICANT: O'Brien, Timothy J.
 APPLICANT: Underwood, Lowell J.
 TITLE OF INVENTION: Transmembrane Serine Protease Overexpressed
 TITLE OF INVENTION: In Ovarian Carcinoma and Uses Thereof
 FILE REFERENCE: D6192CIP
 CURRENT APPLICATION NUMBER: US/09/518,046
 CURRENT FILING DATE: 2000-03-02
 EARLIER APPLICATION NUMBER: 09/261,415
 EARLIER FILING DATE: 1999-03-03
 NUMBER OF SEQ ID NOS: 153
 SEQ ID NO 3
 LENGTH: 2544
 TYPE: DNA
 ORGANISM: Homo sapiens
 FEATURE:
 NAME/KEY: CDS
 OTHER INFORMATION: entire cDNA sequence of TADG-12 variant gen
 -09-518-046-3

1091	GTGGGCTAGTGTTCCTCGTGGCAATCCAGCCCCATCCCCACTCTGGTGGAGAGATGTC	Db
276	IleIleGluPheAsnProMetTyr---ProLys-----AspAsnAspIleAlaLeuMet	Qy
1151	-----FACCACGACGAAAGTACACGCCAAGAGGCTGGGCAATGATCGGCCCTTATG	Db
293	LysLeuGlnPheProLeuThrPheSerGlyThrValArgProIleCysLeuProPhePhe	Qy
1202	AGCTGGCCGGGCACCTCAGTTCATGAATAATGATCCAGCCTGTGTGCTTGCCCACTCT	Db
313	AspGluGluLeuThrProAlaThrProLeuTyrIleIleGlyTyrGlyPheThrLysGln	Qy
1262	GAAGAGAACTTCCCCGATGGAAAAGTGTGCTGCAGCTCAGATGGGGGGCCACA	Db
333	AsnGlyGlyLysMetSerAspIleLeuLeuGlnAlaSerValGlnValIleAspSerThr	Qy

```

b 1319 GATGAGGTGACCCCTCCCTGTCCTGAAACACCGGCGCTCCCTTGATTTCCACAAG 1378
      :::::||||| :::::||||| :::::||||| :::::||||| :::::|||||
y 353 ArgCysAsnAlaAspAspAlaTyrGlnGlyValThrGluLysMetMetCysAlaGly 372
      :::::||||| :::::||||| :::::||||| :::::||||| :::::|||||
b 1379 ATCTGCAACACACAGGACGTTAGCGTGGCATCATCTCCCTCCATGCTCTGGCGGGC 1438
      :::::||||| :::::||||| :::::||||| :::::||||| :::::|||||
y 373 IleProGluGlyValAspThrCysGlnGlyAspSerGlyGlyProLeuMetTyrGln 392
      :::::||||| :::::||||| :::::||||| :::::||||| :::::|||||
b 1439 TACCTGACCGGTGGTGGACAGTGCACGGGGACAGCGGGGGCCCTGGTGTGTCAA 1498
      :::::||||| :::::||||| :::::||||| :::::||||| :::::|||||
y 393 SerAspGln--TyrHisValValGlyValSerThrGlyGlyCysGlyGlyPro 411
      :::::||||| :::::||||| :::::||||| :::::||||| :::::|||||
b 1499 GAGAGAGGCTGTGGAAATGTAGTGGAGCAGCAGCTTTGCGATCGGCTCCGAGAGTG 1558
      :::::||||| :::::||||| :::::||||| :::::||||| :::::|||||
y 412 SerThrProGlyValTyrThrLysValSerAlaTyrLeuAsnTrpIleTyrAsnValTrp 431
      :::::||||| :::::||||| :::::||||| :::::||||| :::::|||||
o 1559 AACAAAGCTGGGGTGTACACCCGTGTACCTCTCTGCTGACTGGATCCACGAGCAGATG 1618
      :::::||||| :::::||||| :::::||||| :::::||||| :::::|||||
y 432 LysAlaGluLeu 435
      :::::|||||
b 1619 GAGAGAGACCTA 1630
      :::::|||||

RESULT 9
S-09-261-416-1
Sequence 1, Application US/09261416A
Patent No. 6291663
GENERAL INFORMATION:
APPLICANT: O'Brien, Timothy J.
APPLICANT: Underwood, Lowell J.
TITLE OF INVENTION: TADG-12: A No. 6291663el Transmembrane Serine Protease
TITLE OF INVENTION: Overexpressed in Ovarian Carcinoma
FILE REFERENCE: D6192
CURRENT APPLICATION NUMBER: US/09/261,416A
CURRENT FILING DATE: 1999-03-03
NUMBER OF SEQ ID NOS: 14
SEQ ID NO 1
LENGTH: 2416
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
LOCATION: 144..1511
OTHER INFORMATION: CDS
S-09-261-416-1

Alignment Scores:
red. No.: 2,1e-57 Length: 2416
core: 621.50 Matches: 158
Percent Similarity: 51.57% Conservative: 72
Local Similarity: 35.43% Mismatches: 150
Query Match: 26.54% Indels: 66
B: 3 Gaps: 18

S-09-607-745-2 (1-435) x US-09-261-416-1 (1-2416)
22 IleProMetClnThrPheArgLysValGlyIleProIleIleIleAlaLeuSerLeu 41
      :::::||||| :::::||||| :::::||||| :::::||||| :::::|||||
b 273 CTGCCATTGAAGTTTTCCTCCCAATCATGCTCATGTTGGGATCATGTCATG 323
      :::::||||| :::::||||| :::::||||| :::::||||| :::::|||||
y 42 AlaSerIleIleIleValValValLeuIleLysValIleLeuAsp-----LysTyr 58
      :::::||||| :::::||||| :::::||||| :::::||||| :::::|||||
b 324 -----ATATTACATCGGCATGCTGCTGGCATCCATTCCTGACTGCTCAGGGAAGTAC 377
      :::::||||| :::::||||| :::::||||| :::::||||| :::::|||||
y 59 -----TyrPheLeuCysGlyGlnProLeuHisPheIleProArgLysGlnLeu 74
      :::::||||| :::::||||| :::::||||| :::::||||| :::::|||||
b 378 AGATGCTGCTCATCTTAAGTGT-----ATCGAGCTGATACTCGA-----419
      :::::||||| :::::||||| :::::||||| :::::||||| :::::|||||
y 75 CysAspGlyGluLeuAspCysProLeuGlyGluAspGluIleHisCysValLysSerPhe 94
      :::::||||| :::::||||| :::::||||| :::::||||| :::::|||||
b 420 TGTGACGGAGTCTCGGATTGTCAAGACGGGGAGGAGGAGTACCGCTGT-----467
      :::::||||| :::::||||| :::::||||| :::::||||| :::::|||||
y 95 ProGluGlyProAlaValAlaValArgLeuSerLysAspArgSerThrLeuGlnValLeu 114
      :::::||||| :::::||||| :::::||||| :::::||||| :::::|||||

```

```

Db 468 -----GTCCGGTGGGTGGTGCAGAAATCCCTGCTCCAGGTGTTCC 506
QY 115 AspSerAlaThrGlyAsnTrpPheSerAlaCysPheAspAsnPheThrGluAlaLeuAla 134
      :::::||||| :::::||||| :::::||||| :::::||||| :::::|||||
Db 507 ACAGCTGCT-----TCGTGGAAGACCATGTGCTCGATGACTGGAAGGTCTACTACGCA 560
      :::::||||| :::::||||| :::::||||| :::::||||| :::::|||||
QY 135 GluThrAlaCysArgGlnMetGlyTyrSer-----145
      :::::||||| :::::||||| :::::||||| :::::||||| :::::|||||
Db 561 AATGTTGCTGTGCCCACTGGGTTTCCCAAGTATGTGAGTTCAGATAACCTCAGAGTG 620
      :::::||||| :::::||||| :::::||||| :::::||||| :::::|||||
QY 146 -----LysProThrPheArgAlaValGluIleGlyProAspGlnAspLeuAspVal 162
      :::::||||| :::::||||| :::::||||| :::::||||| :::::|||||
Db 621 AGCTCGCTGGAGGGCAGTTCGGGAGGAGTTGTGTCCATCATGATCCTCTTCCAGAT 680
      :::::||||| :::::||||| :::::||||| :::::||||| :::::|||||
QY 163 ValGluIleThrGluAsnSerGlnGluLeuArgMetArgAsnSerSerGlyProCysLeu 182
      :::::||||| :::::||||| :::::||||| :::::||||| :::::|||||
Db 681 GACAAAGGTGACTGCATTACACCACTCAGTATATGTGAGGAGGGA-----TGTGCC 731
      :::::||||| :::::||||| :::::||||| :::::||||| :::::|||||
QY 183 SerGlySerLeuValSerLeuHisCysLeuAlaCysGly-----LysSerLeuLysThrPro 201
      :::::||||| :::::||||| :::::||||| :::::||||| :::::|||||
Db 732 TCTGGCCACGTGTTACCTTGCAGTGCACAGCTGTGTGCATGAGGAGGGGTACAGTCA 791
      :::::||||| :::::||||| :::::||||| :::::||||| :::::|||||
QY 202 ArgValValGlyGlyGluGluAlaSerValAspSerTrpProTrpGlnValSerIleGln 221
      :::::||||| :::::||||| :::::||||| :::::||||| :::::|||||
Db 792 CGCATCGTGGTGGAAACATGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 851
      :::::||||| :::::||||| :::::||||| :::::||||| :::::|||||
QY 222 TyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValThrAla 241
      :::::||||| :::::||||| :::::||||| :::::||||| :::::|||||
Db 852 TTCCAGGGGTACCACTGTGCGGGGCTCTGTATCATCAGCCCTCTGGATCATCACTGCT 911
      :::::||||| :::::||||| :::::||||| :::::||||| :::::|||||
QY 242 AlaHisCysPheArgLysHisThrAspValPhe-----AsnTrpLysValArgAla 258
      :::::||||| :::::||||| :::::||||| :::::||||| :::::|||||
Db 912 GCACACTGTGTT-----TATGACTTGTATCTCCCAAGTCATGACCATCCAGGTG 962
      :::::||||| :::::||||| :::::||||| :::::||||| :::::|||||
QY 259 GlySerAspLysLeu-----GlySerPheProSerLeuAlaValAlaLysIleIle 276
      :::::||||| :::::||||| :::::||||| :::::||||| :::::|||||
Db 963 GGTCTAGTTTCTGTTGGACAATCCAGCCCATCCACTTGTGTGGAGAGATGTT--- 1019
      :::::||||| :::::||||| :::::||||| :::::||||| :::::|||||
QY 277 IleGluPheAsnProMetTyr---ProLys-----AspAsnAspIleAlaLeuMetLys 293
      :::::||||| :::::||||| :::::||||| :::::||||| :::::|||||
Db 1020 -----TACCACAGCAAGTACAAAGCCAAAGAGCGTGGGCAATGACATCGCCTTATGAAG 1073
      :::::||||| :::::||||| :::::||||| :::::||||| :::::|||||
QY 294 LeuGlnPheProLeuThrPheSerGlyThrValArgProIleCysLeuProPheAsp 313
      :::::||||| :::::||||| :::::||||| :::::||||| :::::|||||
Db 1074 CTGGCGGGGCCACTCACGTTCAATGAATGATGATCCAGCTGTGTGCTGCCCACTCTGAA 1133
      :::::||||| :::::||||| :::::||||| :::::||||| :::::|||||
QY 314 GluGluLeuThrProAlaThrProLeuTrpIleIleGlyTrpGlyPheThrLysGlnAsn 333
      :::::||||| :::::||||| :::::||||| :::::||||| :::::|||||
Db 1134 GAGAACTTCCCGATGGAAAAGTGTCTGGACGTAGGATGGGGGCCACA---GAGGAT 1190
      :::::||||| :::::||||| :::::||||| :::::||||| :::::|||||
QY 334 GlyGlyLysMetSerAspIleLeuLeuGlnAlaSerValGlnValIleAspSerThrArg 353
      :::::||||| :::::||||| :::::||||| :::::||||| :::::|||||
Db 1191 GGAGGTGACGCTCCCTGCTCTGAACACCGCGCGCTCCCTTGTGTTTCCAAACAAGAT 1250
      :::::||||| :::::||||| :::::||||| :::::||||| :::::|||||
QY 354 -CysAsnAlaAspAspAlaTyrGlnGlyGluValThrGluLysMetMetCysAlaGly 373
      :::::||||| :::::||||| :::::||||| :::::||||| :::::|||||
Db 1251 CTGCAACCAACAGGAGCGTGTACGGTGGCATCATCTCCCTCCATGCTCTGCGGGGCTA 1310
      :::::||||| :::::||||| :::::||||| :::::||||| :::::|||||
QY 373 eProGluGlyGlyValAspThr--CysGlnGlyAspSerGlyGlyProLeuMetTyrGln 392
      :::::||||| :::::||||| :::::||||| :::::||||| :::::|||||
Db 1311 CTGACGGGTGGCTTGGAAAGTGTGACAGCTGCCAGGGGGACAGCGGGGGCCCTGGTGTGTCAA 1370
      :::::||||| :::::||||| :::::||||| :::::||||| :::::|||||
QY 393 SerAspGln---TyrHisValValGlyIleValSerTrpGlyTyrGlyCysGlyGlyPro 411
      :::::||||| :::::||||| :::::||||| :::::||||| :::::|||||
Db 1371 GAGAGAGGCTGTGGAAGTGTAGTGGAGCGCACCACTTTGGCATCGGCTCGCAGACGTTG 1430
      :::::||||| :::::||||| :::::||||| :::::||||| :::::|||||
QY 412 SerThrProGlyValTyrThrLysValSerAlaTyrLeuAsnTrpIleTyrAsnValTrp 431
      :::::||||| :::::||||| :::::||||| :::::||||| :::::|||||
Db 1431 AACAAAGCTGGGGTGTACACCGGTGTCACTCTCTCTGCTGCTGATCCAGCAGATG 1490
      :::::||||| :::::||||| :::::||||| :::::||||| :::::|||||
QY 432 LysAlaGluLeu 435
      :::::|||||
Db 1491 GAGAGAGACCTA 1502
      :::::|||||

```

QY	203	ValValGlyGlyGluGluAlaSerValAspSerTrpProTrpGlnValSerIleGlnTyr	222
Db	732	APTCGTGGGAGCCGGGACACCACTGTGGCCCGGTGGCCCGTGGCACTGACCTTCGCTAT	791
QY	223	AspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThrAlaAla	242
Db	792	GATGAGACACACTCTGTGGGGATCCCTGCTCTCCGGGGACTGGGTGTGACAGCCGCC	851
QY	243	HisCysPheArgLysHisThrAspValPheAsn---TrpLysValArgAlaGlySerAsp	261
Db	852	CACCTGCTCCCGAGCGGGAACCGGGTCTGTCCCGATGCGAGTGTTCGCGTGCCTGT	911
QY	262	LysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleIleGluPheAsnPro	281
Db	912	GCCACAGGCCTCTCCCAACGGTCTGCAGCTGGGGGTGCAGCGCTGTGTCTACCA	971
QY	282	MetTyrPro-----LysAspAsnAspIleAlaLeuMetLys	293
Db	972	TATCTTCCTTCGGGACCCCAACAGCAGGAGGACAGCAACGATATTGCCCTGGTCCAC	1031
QY	294	LeuGlnPheProLeuThrPheSerGlyThrValArgProIleCysLeuProPhePheAsp	313
Db	1032	CTCTCCAGTCCCTGCCTCCACAGATAATCCAGCCGTGTGTGCTCCAGTGCCTGCCG	1091
QY	314	GluGluLeuThrProAlaThrProLeuTrpIleIleGlyTrpGlyPheThrLysGlnAsn	333
Db	1092	CAGGCCCTGTGTGATGGCAAGATCTGTACCGTGACCGGGCTGGGGCAACAG---CAGTAC	1148
QY	334	GlyGlyLysMetSerAspIleLeuGlnAlaSerValGlnValIleAspSerThrArg	353
Db	1149	TATGGCCAAACAGCCGGGTACTCCAGGAGGCTCGAGTCCCAATATACCAATGATGTC	1208
QY	354	CysAsnAlaAspAlaTrpGlnGlyGluValThrGluLysMetCysAlaGlyIle	373
Db	1209	TGCAATGGCGCTGACTTCTATGGAAACACAGATCAAGCCCAAGATGTCTGTGCTGGCTAC	1268
QY	374	ProGluGlyGlyValAspThrCysGlnGlyAspSerGlyGlyProLeuMetTyrGln---	392
Db	1269	CCGAGGTGGCAATTCATGCTGCCAGGGCGACAGCGGTGTGCTCCCTTGTGTGTGAGGAC	1328
QY	393	-----SerAspGlnTrpHisValValGlyIleValSerTrpGlyTyrGlyCys	408
Db	1329	AGCATCTCTCGACGCCACGTGTGGCGCTGTGTGGCATTTGTGATTTGGGCACTGGCTGT	1388
QY	409	GlyGlyProSerThrProGlyValTyrThrLysValSerAlaTyrIleuAsnTrpIleTyr	428
Db	1389	GCCTGGCCAGCAGCGCGCTTACCCAAAGTCACTGCTCCGGGAGTGGATCTTC	1448
QY	429	AsnValTrpLys 432	
Db	1449	CAGGCCATAAAG 1460	
RESULT 11			
US-09-861-966-188			
; Sequence 188, Application US/09861966			
; Patent No. 6518028			
; GENERAL INFORMATION:			
; APPLICANT: O'Brien, Timothy J.			
; TITLE OF INVENTION: Compositions and Methods for the Early Diagnosis of			
; FILE REFERENCE: D6223CIP-A/Div			
; CURRENT FILING DATE: 2001-05-21			
; PRIOR APPLICATION NUMBER: 09/510,738			
; PRIOR FILING DATE: 2000-02-22			
; NUMBER OF SEQ ID NOS: 188			
; SEQ ID NO 188			
; LENGTH: 1783			
; TYPE: DNA			
; ORGANISM: Homo sapiens			
; FEATURE:			
; OTHER INFORMATION: full length cDNA of hepsin			
US-09-861-966-188			

Alignment Scores:
 Pred. No.: 1,366-53 Length: 1783
 Core: 584.50 Matches: 140
 Percent Similarity: 45.05% Conservative: 60
 Best Local Similarity: 31.53% Mismatches: 161
 Query Match: 24.96% Indels: 83
 Gaps: 4

US-09-607-745-2 (1-435) x US-09-861-966-188 (1-1783)

19 LysProArgIleProMetGluThrPheArgLysValGlyLeuProIleIleAlaLeu 38
 288 AGACCAAGTGGCAGCTCTCACT-----GCGGGACCTCTACTTCTGACAGCC 338
 39 LeuSerLeuAlaSerIleIleValValLeuIleLysValIleLeuAspLysTyr 58
 339 ATCGGGGGCATCTCTGGCCATTGGCTGTCTCTCAGG-----380
 59 TyrPheLeuCysGlyGlnProLeuHisPheIleProArgLysGlnLeuCysAspGlyGlu 78
 380 -----380
 79 LeuAspCysProLeuGlyGluAspGluGluHisCysValLysSerPheProGluGlyPro 98
 381 -----ACTGACCGAGG-----CCG 395
 99 AlaValAlaValArgLeuSerLysAspArgSerThrLeuGlnValLeuAspSerAlaThr 118
 396 CTGTACCCAGTCAGGTGAGTCTGGGAGCTCGGCTCATGCTCTTGACAGAGCGAA 455
 119 GlyAsnTrpPheSerLacCysPheAspAsnPhThrGluAlaLeuAlaGluThrAlaCys 138
 456 GGGACGTGGCGGTGTGTCTCTCTCTCTCCAGCCAGCGGTAGCGGACCTCAGTGC 515
 139 ArgGlnMetGlyTyrSerSerLysProThrPheArgAlaValGluIleGlyProAspGln 158
 516 GAGGAGATGGCTTC-----CTCAGGGCACTG-----ACCCACTCC 551
 159 AspLeuAspValValGluIleThrGluAsn-----168
 552 GAGCTGGAGTGGCAAGCGGGCGGCAATGGCAGCTCGGGCTTCTCTGTGGACGAG 611
 169 -----SerGlnGluLeuArgMetArgAsnSerSerGlyProCysLeuSer 183
 612 GGGAGGCTGCCACACCCAGAGGCTCTGAGGTATCTCCGTGTGTGATGCCCCAGA 671
 184 GlySerLeuValSerLeuHisCysLeuAlaCysGly---LysSerLeuLysThrProArg 202
 672 GCGCGTTCTTGGCGGCATCTGCCAAGACTGTGGCGCAGGAAGTGCCTGGGACCGC 731
 203 ValValGlyGlyGluGluAlaSerValAspSerTrpProTrpGlnValSerIleGlnTyr 222
 732 ATCGTGGAGCGCGGACACCACTTGGCGGTGGCGTGGCAAGTCAAGCTTCGCTAT 791
 223 AspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThrAlaAla 242
 792 GATGAGACACACCTCTGTGGGGATCTCTCTCCGGGACTGGGTGTGACAGCGCC 851
 243 HisCysPheArgLysHisThrAspValPheAsn---TrpLysValArgAlaGlySerAsp 261
 852 CATGCTTCCGGAGCGGACCGGCTCTGTCGATGGCGAGTGTTCGGTGGCGGTG 911
 262 LysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleGluPheAsnPro 281
 912 GCCAGGCTCTCCACCGTCTGACCTGGGGGTGGCAGCTGTGGTCTACCGGGGGC 971
 282 MetTyrPro-----LysAspAsnAspIleAlaLeuMetLys 293
 972 TATCTTCCCTTCGGGACCCCAACAGCGAGGAGAACAGCATATGCTCTGCTCCAC 1031
 294 LeuGlnPheProLeuThrPheSerGlyThrValArgProIleCysLeuProPheAsp 313

1032 CTCTCAGTCCCTCCCTCCACATATACATCCAGCTGTGTGCTCCAGTCCCGGC 1091
 314 GluGluLeuThrProAlaThrProLeuTrpIleIleGlyTrpGlyPheThrLysGlnAsn 333
 1092 CAGGCGCTGGTGGTGAAGATCTGTACCGTGAACGGCTGGGCAACAGC---CAGTAC 1148
 334 GlyGlyLysMetSerAspIleLeuLeuGlnAlaSerValGlnValIleAspSerThrArg 353
 1149 TATGGCCAAAGCGCGGGGTACTCCAGGAGGTCCAGTCCCATATACAGCATGATC 1208
 354 CysAsnAlaAspAlaTyrGlnGlyGluValThrGluLysMetMetCysAlaGlyIle 373
 1209 TCATATGGCGTGAATCTTATGGAACACGATCAAGCCAAAGATGTTCTGTCTGGCTAC 1268
 374 ProGluGlyGlyValAspThrCysGlnGlyAspSerGlyGlyProLeuMetTyrGln--- 392
 1269 CCGAGGTGGCATTCATGCTGCCAGCGCAGCGGTGGTCCCTTTGTGTGAGGAC 1328
 393 -----SerAspGlnTrpHisValValGlyIleValSerTrpGlyTyrGlyCys 408
 1329 AGCATCTCTCGACGCCAGTGGCGGTGTGTGCATTTGAGTTGGGCACTGGCTGT 1388
 409 GlyGlyProSerThrProGlyValTyrThrLysValSerAlaTyrLeuAsnTrpIleTyr 428
 1389 GCCTGGGCCAAGACCGGCGCTCTACACCAAGTCAGTGACTTCGGGAGTGATCTTC 1448
 429 AsnValTrpLys 432
 1449 CAGGCCATAAAG 1460

RESULT 12
 US-09-742-703-3
 ; Sequence 3, Application US/09742703
 ; Patent No. 6423543
 ; GENERAL INFORMATION:
 ; APPLICANT: Patrick Allen Marcotte
 ; APPLICANT: Lex M. Cowert
 ; TITLE OF INVENTION: ANTISENSE MODULATION OF HEPSEIN EXPRESSION
 ; FILE REFERENCE: RTS-0090
 ; CURRENT APPLICATION NUMBER: US/09/742,703
 ; CURRENT FILING DATE: 2000-12-20
 ; NUMBER OF SEQ ID NOS: 49
 ; SEQ ID NO 3
 ; LENGTH: 2363
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (826)..(2079)
 US-09-742-703-3

Alignment Scores:
 Pred. No.: 2,116-53 Length: 2363
 Score: 584.50 Matches: 140
 Percent Similarity: 45.05% Conservative: 60
 Best Local Similarity: 31.53% Mismatches: 161
 Query Match: 24.96% Indels: 83
 Gaps: 4

US-09-607-745-2 (1-435) x US-09-742-703-3 (1-2363)

19 LysProArgIleProMetGluThrPheArgLysValGlyLeuProIleIleAlaLeu 38
 868 AGACCAAGTGGCAGCTCTCACT-----GCGGGACCTCTACTTCTGACAGCC 918
 39 LeuSerLeuAlaSerIleIleValValLeuIleLysValIleLeuAspLysTyr 58
 919 ATCGGGGGCATCTCTGGCCATTGGCTGTCTCTCAGG-----960
 59 TyrPheLeuCysGlyGlnProLeuHisPheIleProArgLysGlnLeuCysAspGlyGlu 78
 960 -----960

79 LeuAspCysProLeuGlyGluAspGluHisCysVallysserPheProGluGlyPro 98
961 -----AGTGACAGAG-----CCG 975
99 AlavalalaValargLeuSerLysAspArgSerThrLeuGlnValLeuAspSerAlaThr 118
976 CTGTACCAGTCAGGTCAGCTCTGCGAGAGCTCGGCTCATGTGTTTGACAAGCGAA 1035
119 GlyAsnTrpPheSerAlaCysPheAspAsnPheThrGluAlaLeuAlaGluThrAlaCys 138
1036 GGGACGTGGGGTGTGTCTCTCGCGCTCCACGCCAGGTAGCGGACTCAGCTGC 1095
139 ArgGlnMetGlyTyrSerSerLysProThrPheArgAlaValGluIleGlyProAspGln 158
1096 GAGGAGATGGCTTC-----CTCAGGCGACTG-----ACCCACTCC 1131
159 AspLeuAspValValGluIleThrGluAsn-----168
1132 GAGCTGACGTGCGAAGCGCGGGCCCAATGGCAGCTCGGGCTCTCTGTGTGGAGCAG 1191
169 -----SerGlnGluLeuArgMetArgAsnSerSerGlyProCysLeuSer 183
1192 GGGAGGCTGCCCCACACCCAGAGGCTGCTGGAGGTCTCTCCGTGTGATGGCCCCAGA 1251
184 GlySerLeuValSerLeuHisCysLeuAlaCysGly---LysSerLeuLysThrProArg 202
1252 GGGCGTTTCTTGGCGGCTCTGCTCCAGACTGTGTGGCGCAGGAAGCTCCCGTGGACCGC 1311
203 ValValGlyGlyGluGluAlaSerValAspSerTrpProGlnValSerIleGlnTyr 222
1312 ATCTGGAGCGCGGACACCCAGCTGTGGCGGTGGCGGTGGAGTGTCTTCGCTAT 1371
223 AspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThrAlaAla 242
1372 GATGAGCACACCTCTGTGGGGATCTCTGCTCCGGGAGCTGGGTGCTGACAGCGGCC 1431
243 HisCysPheArgLysHisThrAspValPheAsn---TrpLysValArgAlaGlySerAsp 261
1432 CACTGCTTCCGGAGCGGAACCGGCTCTGCTCCGAGTGGCGAGTGTTCGGTGGCGGTG 1491
262 LysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleGluPheAsnPro 281
1492 GCCAGCGCTCTCCCCAGCTGTCTGAGTGGGTGGGTGAGGCTGTGGTACCAGGGGGC 1551
282 MetTyrPro-----LysAspAsnAspIleAlaLeuMetLys 293
1552 TATCTTCCCTTCGGACCCCAACAGCGAGAGACAGCAAGTATATGCTGTGTCCAC 1611
294 LeuGlnPheProLeuThrPheSerGlyThrValArgProIleCysLeuProPhePheAsp 313
1612 CTCTCCAGTCCCTGCCCCCTCACAGAAATACATCAGGCTGTGTGCTCCAGCTGCCGGC 1671
314 GluGluLeuThrProAlaThrProLeuTrpIleIleGlyTrpGlyPheThrLysGlnAsn 333
1672 CAGGCTGTGTGGATGAGTGAAGATCTGTACGTGAGGGCTGGGCGCAACAG---CAGTAC 1728
334 GlyGlyLysMetSerAspIleLeuLeuGlnAlaSerValGlnValIleAspSerThrArg 353
1729 TATGGCCAAACAGGCGGGGTACTCCAGGAGGCTCGAGTCCCAATATCAGCAATGATGTC 1788
354 CysAsnAlaAspAspAlaTyrGlnGlyGluValThrGluLysMetMetCysAlaGlyIle 373
1789 TGCAATGCGGTGACTTCTATGGAAACAGATCAAGCCCAAGATGTTCTGTGTGGGTAC 1848
374 ProGluGlyGlyValAspThrCysGlnGlyAspSerGlyGlyProLeuMetTyrGln--- 392
1849 CCCAGGGTGCATGATGCTCTGCGAGGGCAGCGGTGTGCTCTTGTGTGTGGAGAC 1908
393 -----SerAspGlnTrpHisValValGlyIleValSerTrpGlyTyrGlyCys 408
1909 AGCATCTCTCGGACCGCAGTGTGGCGGTGTGTGGCATTGTGAGTTGGGGACTGTGTGT 1968
409 GlyGlyProSerThrProGlyValTyrThrLysValSerAlaTyrLeuAsnTrpIleTyr 428

Db 1969 GCCTGGCCAGAGCCAGCGCTCTACACAAAGTCAGTCACTTCGGGGAGTGATCTTC 2028
Qy 429 AsnValTrpLys 432
Db 2029 CAGGCCATTAAG 2040
RESULT 13
US-09-820-002-1
; Sequence 1, Application US/09820002
; Patent No. 6482630
; GENERAL INFORMATION:
; APPLICANT: Gan, Weiniu
; APPLICANT: Ye, Jane
; APPLICANT: DiFrancesco, Valentina
; APPLICANT: Beasley, Ellen
; TITLE OF INVENTION: ISOLATED HUMAN PROTEASE PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES, ENCODING HUMAN PROTEASE PROTEINS, AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: CL001194
; CURRENT APPLICATION NUMBER: US/09/820,002
; CURRENT FILING DATE: 2001-03-29
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1615
; TYPE: DNA
; ORGANISM: HUMAN
US-09-820-002-1
Alignment Scores:
Pred. No.: 3.6e-53 Length: 1615
Score: 580.00 Matches: 132
Percent Similarity: 44.06% Conservative: 57
Best Local Similarity: 30.77% Mismatches: 146
Query Match: 24.77% Indels: 94
DB: 4 Gaps: 9
US-09-607-745-2 (1-435) x US-09-820-002-1 (1-1615)
Qy 19 LysProArgIleProMetGluThrPheArgLysValGlyIleProIleIleAlaLeu 38
Db 220 AGACCAAGGTGGCAGCTCTCACT-----GCGGGAGCCCTGCTACTTCTGACAGCC 270
Qy 39 LeuSerLeuAlaSerIleIleIleValValLeuIleLysValIleLeuAspLysTyr 58
Db 271 ATCGGGGGCATCTCTGGCCATGTGGCTGTCTCTCAG-----312
Qy 59 TyrPheLeuCysGlyGlnProLeuHisPheIleProArgLysGlnLeuCysAspGlyGlu 78
Db 312 -----312
Qy 79 LeuAspCysProLeuGlyGluAspGluGluHisCysVallysserPheProGluGlyPro 98
Db 313 -----AGTGACAGAG-----CCG 327
Qy 99 AlavalalaValargLeuSerLysAspArgSerThrLeuGlnValLeuAspSerAlaThr 118
Db 328 CTGTACCAGTCAGGTCAGCTCTGCGAGGCTCGGCTCATGTGTTTGACAAGCGAA 387
Qy 119 GlyAsnTrpPheSerAlaCysPheAspAsnPheThrGluAlaLeuAlaGluThrAlaCys 138
Db 388 GGGACGTGGGGTGTGTCTCTCGCGCTCCACGCCAGGTAGCGGACTCAGCTGC 447
Qy 139 ArgGlnMetGlyTyrSerSerLysProThrPheArgAlaValGluIleGlyProAspGln 158
Db 448 GAGGAGATGGCTTCTCTCAGT-----468
Qy 159 AspLeuAspValValGluIleThrGluAsnSerGlnGluLeuArgMetArgAsnSer 178
Db 468 -----468
Qy 179 GlyProCysLeuSerGlySerLeuValSerLeuHisCysLeuAlaCysGly---LysSer 197

db 469 ---GATTGCCCCAGAGCCGTTCTTGGCCCGCCATCTGCCAAGACTGTGGCCGAGGAAG 525
y 198 LeuLysThrProArgValValGlyGluGluAlaSerValAspSerTrpProTyrGln 217
b 526 CTGCGCGTGGACCGCATCGTGGGAGGCGGAGACACCGAGTGTGGCGGTGGCGGTGCA 585
y 218 ValSerIleGlnTyrAspLysGlnHisValCysGlySerIleLeuAspProHisTyr 237
b 586 GTACAGCCTTGGCTATGATGAGACACCTCTGTGGGGATCCCTGCTCTCGGGGACTGG 645
y 238 ValLeuThrAlaAlaHisCysPheArgLysHisThrAspValPheAsn---TyrLysVal 256
b 646 GTGCTGACACCGCCACTGCTTCCGAGCGGACCGGTCCTGTCGCCGATGGCGAGTG 705
y 257 ArgAlaGlySerAspLysLeuGlySerPheProSerLeuAlaValAlaLysIle 276
b 706 TTTGCGGTGCGGCGCCAGGCTCTCCACAGGTCTGCAGCTGGGGGTGCAGGCTGTG 765
y 277 IleGluPheAsnProMetTyrPro-----LysAspAsnAsp 288
b 766 GTCTACACCGGGGCTATCTTCCCTTTCCGACCCCAACACGCGAGGAGAACAGCAACGAT 825
y 289 IleAlaLeuMetLysLeuGlnPheProLeuThrPheSerGlyThrValArgProIleCys 308
b 826 ATTGCGCTGGTCCACTCTCCAGTCCCTGCTCCCTACAGATACATCCAGCCTGTGTC 885
y 309 LeuProPhePheAspGluGluLeuThrProAlaThrProLeuTyrPheIleGlyTyrGly 328
b 886 CTCCAGCTGCGGCGCCAGGCTGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 945
y 329 PheThrLysGlnAsnGlyGlyLysMetSerAspIleLeuLeuGlnAlaSerValGlnVal 348
b 946 AACAGC---CAGTACTATGGCCACAGCCGGGGTACTCCAGAGGCTCGAGTCCCAT 1002
y 349 IleAspSerThrArgCysAsnAlaAspAlaTyrGlnGlyGluValThrGluLysMet 368
b 1003 ATCAGCAATGATGCTGCAATGGCGCTGACTTCTATGGAAACAGATCAAGCCGAGATG 1062
y 369 MetCysAlaGlyIleProGluGlyGlyValAspThrCysGlnGlyAspSerGlyPro 388
b 1063 TTCTGTGCTGCTACCGCGAGGTGGATGATGCTGCTGCGAGGCGACAGCGGTGGTCCC 1122
y 389 LeuMetTyrGln-----SerAspGlnTyrHisValValGlyIleValSer 403
b 1123 TTTGTGTGAGGACAGCATCTTCGAGCCACGCTTGGCGGCTGTGGCGATGTGAGT 1182
y 404 TrpGlyTyrGlyCysGlyGlyProSerThrProGlyValTyrThrLysValSerAlaTyr 423
b 1183 TGGGCACTGGCTGTGCGCTGGCCCGAGAGCCAGGCGTCTACACCAAGTCAGTCACTTC 1242
y 424 LeuAsnTrpIleTyrAsnValTyrLys 432
b 1243 CGGAGTGGATCTTCCAGGCCATAAAG 1269

RESULT 14

US-09-000-846-1
Sequence 1, Application US/09000846
Patent No. 5981830
GENERAL INFORMATION:
APPLICANT: WU, QINGYU
APPLICANT: SADLER, JASPER
TITLE OF INVENTION: KNOCKOUT MICE AND THEIR PROGENY WITH
TITLE OF INVENTION: DISRUPTED SERINE PROTEASE GENES
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: MILLEN, WHITE, ZELANO & BRANTIGAN, P.C.
STREET: 2200 CLARENDON BLVD. SUITE 1400
CITY: ARLINGTON
STATE: VA
COUNTRY: US
ZIP: 22201
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/000,846
FILING DATE: 30-DEC-1997
CLASSIFICATION: 800
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/866,058
FILING DATE: 30-MAY-1997
ATTORNEY/AGENT INFORMATION:
NAME: LEBOVITZ, RICHARD M.
REGISTRATION NUMBER: 37,067
REFERENCE/DOCKET NUMBER: BERLX 65P1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-243-6333
TELEFAX: 703-243-6410
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1605 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE: ORGANISM: Mus musculus
FEATURE:
NAME/KEY: CDS
LOCATION: 35..1282
US-09-000-846-1
Alignment Scores:
Pred. No.: 5.88e-53 Length: 1605
Score: 578.00 Matches: 136
Percent Similarity: 47.43% Conservative: 58
Best Local Similarity: 33.25% Mismatches: 152
Query Match: 24.68% Indels: 63
DB: 2 Gaps: 11
US-09-607-745-2 (1-435) x US-09-000-846-1 (1-1605)
Qy 62 CysGlyGlnProLeuHisPheIleProArgLysGln-----LeuCysAspGlyGluLeu 79
Db 97 TGTGGTACCTCTCT-----GTCTGACAGCATTTGGGGCCGCGCTCTGGGCAATTGT 150
Qy 80 AspCysProLeuGlyGlu-AspGluGluHisCysValLysSerPheProGluGlyProAl 99
Db 151 GACATCTACTCAGAGTACCAGGAG-----CCACT 183
Qy 99 aValAlaValArgLeuSerLysAspArgSerThrLeuGlnValLeuAspSerAlaThrGl 119
Db 184 GTACCAAGTGCAGTCCAGTCCAGGGGACTCAGCGTTGCGGTGTTTGACAAGACGGAGGG 243
Qy 119 YAsnTrpPheSerAlaCysPheAspAsnPheThrGluAlaLeuAlaGluThrAlaCysAr 139
Db 244 AACGTGGAGGCTACTGTGCTCTCCTCAGCTCCAAATGCCAGGGTGGCAGGGCTCGGCTGA 303
Qy 139 gGlnMetGlyTyr----- 143
Db 304 GGAGATGGGCTTTCTCAGGCTCTGCGGCACCTCGAGCTGGATGTCCGCGGCGCGC 363
Qy 144 -SerSerLysProThrPheArgAlaValGluIleGly-----ProAspGlnAs 159
Db 364 CACGGGCACATCGGCTTCTTTTTCGTCGACGAGGCGGAGTCCGTCGCTCAGAGGTT 423
Qy 159 pLeuAspValValGluIleThrGluAsnSerGlnGluLeuArgMetArgAsnSerSerGl 179
Db 424 GCTGGATGTCTCTGTATGTGAC----- 448
Qy 179 yProCysLeuSerGlySerLeuValSerLeuHisCysLeuAlaCysGly---LysSerLe 198

```
449 ----TGCTCTAGAGCGGCGATTCTGACTGCGACCTGCGCAAGACTGTGGCGCGAGGAAGCT 504
198 ulysThrProArgValValGlyGluGluAlaSerValAspSerTrpProTrpGlnVa 218
505 GCGGTGGACCGGCGATTGTGGGGGCGGAGCAGAGCTGGGAGGCGGCGGCGGAGGT 564
218 lSerIleGlnTyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpVa 238
565 CAGCTCGGTATGATGGGACCGACCTCTGTGGGGGTCCCTGCTGCTGGGAGCTGGGT 624
238 lLeuThrAlaAlaHisCysPheArgLysHisThrAspValPheAsn---TrpLysValAr 257
625 GCTGACTGCTGCACATGCTTTCAGAGCGGACCGGGTCTCTGCTCGGTGGCGAGTATT 684
257 gAlaGlySerAspLysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIle 277
685 TGCTGGTGTGTAGCCCGGACCTCACCCATGCTGTGCACTGGGGGTTCAGGCTGTGAT 744
277 eGluPheAsnProMetTyrPro-----LysAspAsnAspIle 289
745 CTATCATGGGGGTACCTTCCCTTCGAGACCCCTACTATCGACGAAACAGCAATGACAT 804
289 eAlaLeuMetLysLeuGlnPheProLeuThrPheSerGlyThrValArgProIleCysLe 309
805 TGCTTGTGTCACCTCTCTAGCTCCCTCTCTCAGCAATATACATCCAGCCAGTGTCTCT 864
309 uproPhePheAspGluGluLeuThrProAlaThrProLeuTrpIleIleGlyTrpGlyPh 329
865 CCCTGTGCGGACAGGCGCTGGTGATGGCAAGCTGTGACTGTGACCGCTGGGGTAA 924
329 eThrLysGlnAsnGlyGlyLysMetSerAspIleLeuLeuGlnAlaSerValGlnValIle 349
925 CACA--CAGTTCATGCGCAACAGGCTATGCTGCCAAGAGGCGCGGTGCCATCAT 981
349 eAspSerThrArgCysAsnAlaAspAspAlaTyrGlnGlyGluValThrGluLysMetMe 369
982 AAGCAACGAAGTTTGCAACAGCCCGACTTCTACGGGAATCAGATCAGCCCAAGATGT 1041
369 tCysAlaGlyIleProGluGlyGlyValAspThrCysGlnGlyAspSerGlyGlyProLe 389
1042 CTGTGCTGGCTATCTCTGAGGGTGGCAATGATGCTGCCAGGCGGACAGTGGAGGCCCTT 1101
389 uMetTyrGln-----SerAspGlnTrpHisValValGlyIleValSerTr 404
1102 TGCTGTGAAGACACATCTCTGGGACATCAAGGTGGCGCTATGTGCAATTGTAAGCTG 1161
404 pGlyTyrGlyCysGlyGlyProSerThrProGlyValTyrThrLysValSerAlaTyrLe 424
1162 GGGTACGGGCTGTGCTTTGGCCCGGAGCGAGGTGTACACCAAGTCACTGACTTCGG 1221
424 uAsnTrpIleTyrAsnValTrpLys 432
1222 GGATGGATCTTCAAGGCCATTAAG 1246
```

RESULT 15

```
US-09-370-838-80
; Sequence 80, Application US/09370838
; Patent No. 644425
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Mohamath, Roadon
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOUNDS FOR THERAPY AND DIAGNOSIS OF
; FILE REFERENCE: 210121.475C1
; CURRENT APPLICATION NUMBER: US/09/370,838
; CURRENT FILING DATE: 1999-08-09
; EARLIER APPLICATION NUMBER: US 09/285,323
; EARLIER FILING DATE: 1999-04-02
; NUMBER OF SEQ ID NOS: 289
; SOFTWARE: FastSeq for Windows Version 3.0
```

```
; SEQ ID NO 80
; LENGTH: 1460
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-370-838-80

Alignment Scores:
Pred. No.: 8,37e-53 Length: 1460
Score: 576.00 Matches: 145
Percent Similarity: 50.88% Conservative: 86
Best Local Similarity: 31.94% Mismatches: 161
Query Match: 24,59% Indels: 62
DB: 4 Gaps: 16

US-09-607-745-2 (1-435) x US-09-370-838-80 (1-1460)

Qy 5 SerAspGlnProLeuAsnSerLeuAspValLysProLeuArgLysProArgIleProMet 24
Db 19 GCAGAAAAGAACCTCTTCATTAAGATTAAATGATAGCCAGCAGCTGTA---ACT 75
Qy 25 GluThrPheArgLysValGlyIleProIleIleAlaLeuLeuSerLeuAlaSerIle 44
Db 76 TCGACTTCAAGATTCTCGAAT---CCATATGTAGTATGTTTCATTGCTCGCAGGGGTA 132
Qy 45 IleIleValValLeuIleLysValIleLeuAspLysTyrPheLeuCysGlyGln 64
Db 133 GTGATCTGCGCAGTCCATAGCTACTTGT---TACTTTTAGCTTTTGAT 183
Qy 65 ProLeuHisPheIleProArgLys-----GlnLeuCysAspGlyGluLeuAspCysPro 82
Db 184 CAAAATCTTACTTTTATAGAGCAGTTTCACTCTCTAAATGTGATATATAATAGTCAG 243
Qy 83 Leu-----GlyGluAspGluGluHisCys 90
Db 244 TTTAAATTCACAGCTACACAGGAATACAGGACTTTGAGTGGAAAGATTGAATCTCTGATT 303
Qy 91 ValLysSerPheProGluGly-----ProAlaValAlaValArg 103
Db 304 ACTAAACATTCAAAGAAATCAAAATTAAGAAATCAGTTCATCAGAGCTCATGTTGCCAAA 363
Qy 104 LeuSerLysAsp-----ArgSerThrLeuGlnValLeuAspSerAlaThrGly 119
Db 364 CTGAGGCAATGAGTGTGTGAGAGCGGATGTTCTCATGAATTTCAATTCACATAGA 423
Qy 120 AsnTrpPheSerAlaCysPheAspAsnPheThrGluAlaLeuAlaGluThrAlaCysArg 139
Db 424 AATAACAATGAGCATCAATG-----AAAAGCAGAAATTGAGTCTGTTTACGA 471
Qy 140 GlnMetGlyTyrSerSerLysProThrPheArgAlaValGluIleGlyProAspGlnAsp 159
Db 472 CAATGTCTGAATAACTCT-----GGAAACCTGGAAATAAACCCCT-----TCA 513
Qy 160 LeuAspValValGluIleThrGluAsnSerGlnGluLeuArgMetArgAsn-----Ser 177
Db 514 ACTGAGATAACATCATTACTGACGAGCTGCACCAATTTGGCTTATTATGATGATGGG 573
Qy 178 SerGlyProCysLeuSerGlySerLeuValSerLeuHisCysLeuAlaCysGlyLysSer 197
Db 574 GCGGTCCAGACCTAATA-----ACAA-----ACA 594
Qy 198 LeuLysThrProArgValValGlyGluGluAlaSerValAspSerTrpProTrpGln 217
Db 595 TTGTCTGAGCAGAGAAATCCTTTGGAGGCACTGAGCTGAGGAGGAAGTGCCTGCGGCA 654
Qy 218 ValSerIleGlnTyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrp 237
Db 655 GTCACTGTGGCGCTCAATAATGCCACCATCTGTGGAGGCACCTGATCAATACATGGG 714
Qy 238 ValLeuThrAlaAlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArg 257
Db 715 ATCTTGACAGCAGCTCACTGCTTCAGAGCAACTCTAATCTCGTGTGACTGAGTGGCCACG 774
Qy 258 AlaGlySerAspLysLeuGlySerPheProSerLeu-----AlaValAlaLysIle 275
```

```

775 TCCTGGTATTCACCA-----ACATTTCCCTAACTAGATGAGAGTAGAATAATATTTTA 828
276 IleIleGluPheAsnProMetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGln 295
829 ATTCAATAACAATTAATAAATCTGCAACTCATGAAATGACATTGCACCTTGTGAGACTTTGAG 888
296 PheProLeuThrPheSerGlyThrValArgProIleCysLeuProPhePheAspGluGlu 315
889 AACAGTGTACCTTTTACCAAGATATCCATAGTGTGTCTCCAGCTGCTACCCAGAAT 948
316 LeuThrProAlaThrProLeuThrPheIleGlyThrGlyPheThrLysGlnAsnGlyGly 335
949 ATTCCACCTGGCTCTACTGCTTATGTAAACAGGATGGGGCGCTCAAGAAATATGCTGGCCAC 1008
336 LysMetSerAspIleLeuLeuGlnAlaSerValGlnValIleAspSerThrArgCysAsn 355
1009 ACAGTTCACAGAG--CTAAGGCAAGGACAGTCAAGATAATAAGTAATGATGTATGTAAT 1065
356 AlaAspAlaTyrGlnGlyGluValThrGluLysMetMetCysAlaGlyIleProGlu 375
1066 GCACCAACATAGTTATATATGAGGCCATCTTGTCTGGAATGCTGTGTGGAGTACCTCAA 1125
376 GlyGlyValAspThrCysGlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGln 395
1126 GGTGGAGTGGACGCACTCAGGTGACTCTGGTGGCCCACTAGTACAGAAGACTCACGG 1185
396 -----TrpHisValValGlyIleValSerTrpGlyTyrGlyCysGlyGlyProSerThr 413
1186 CGGCTTTGGTTTATTGTGGGATAGTAAGCTGGGGAGATCAGTGTGGCCTGCCGATAAG 1245
414 ProGlyValTyrThrLysValSerAlaTyrIleuAsnTrpIle 427
1246 CCAGGAGTGTATACTCGAGTCACAGCCTACCTTGACTGGATT 1287
```

Search completed: November 23, 2003, 12:08:53
DB time : 126.359 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

protein - nucleic search, using frame_plus_p2n model

on on: November 23, 2003, 07:54:55 ; Search time 467.909 Seconds
(without alignments)
3038.343 Million cell updates/sec

icle: US-09-607-745-2

effect score: 2342

sequence: 1 MDPDSQPLNSLDVPLRKP.....VYTKVSAYLNWYVWKAEL 435

coring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

sarched: 2169961 seqs, 1634102185 residues

total number of hits satisfying chosen parameters: 4339922

imum DB seq length: 0

aximum DB seq length: 2000000000

ost-processing: Maximum Match 0%

Listing first 45 summaries

ommand line parameters:

MODEL=frame+ p2n.model -DEV=xlh
C=/cg2_1/USPTO.spool/US09607745/runat 21112003 144347 22270/app query.fasta_1.1038
DB=Published Applications NA -QFMT=fastap -SUFFIX=rnpb -MINWATCH=0.1
LOOFCU=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=bloms62
TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100
THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HRPRTZ=500 -MINLEN=0
MAXLEN=200000000 -USER=US09607745 @CGN 1 1 424 @runat 21112003 144347 22270
NCPU=6 -ICPU=3 -NO_MMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100
LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

atabase : Published Applications NA:

1: /cg2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cg2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cg2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cg2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cg2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cg2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cg2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cg2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cg2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
10: /cg2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
11: /cg2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
12: /cg2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
13: /cg2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
14: /cg2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
15: /cg2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
16: /cg2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
17: /cg2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

result No.	Score	Query Match	Length	ID	Description
1	2338	99.8	2038	12	US-10-180-719-18
					Sequence 18, Appl

2	2337	99.8	2081	9	US-09-851-588-7	Sequence 7, Appli
3	2337	99.8	2307	14	US-10-097-340-317	Sequence 317, App
4	2337	99.8	2307	14	US-10-171-311-217	Sequence 217, App
5	2329	99.4	2165	12	US-10-101-510-634	Sequence 634, App
6	2324	99.2	2079	9	US-09-851-588-5	Sequence 5, Appli
7	2324	99.2	2079	11	US-09-776-191-71	Sequence 71, Appl
8	2324	99.2	2079	14	US-10-264-820-22	Sequence 22, Appl
9	2324	99.2	2079	14	US-10-254-289-1	Sequence 1, Appli
10	2319	99.0	2137	11	US-09-776-191-3	Sequence 3, Appli
11	2297.5	98.1	2063	11	US-09-888-257A-2	Sequence 2, Appli
12	2297.5	98.1	2063	11	US-09-946-374-274	Sequence 274, App
13	2297.5	98.1	2063	12	US-10-015-387A-274	Sequence 274, App
14	2297.5	98.1	2063	12	US-10-063-735-111	Sequence 111, App
15	2297.5	98.1	2063	12	US-10-006-130A-274	Sequence 274, App
16	2297.5	98.1	2063	12	US-10-199-672-329	Sequence 329, App
17	2297.5	98.1	2063	12	US-10-006-172A-274	Sequence 274, App
18	2297.5	98.1	2063	12	US-10-187-749-329	Sequence 329, App
19	2297.5	98.1	2063	12	US-10-194-457-329	Sequence 329, App
20	2297.5	98.1	2063	12	US-10-184-642-329	Sequence 329, App
21	2297.5	98.1	2063	12	US-10-196-747-329	Sequence 329, App
22	2297.5	98.1	2063	12	US-10-015-392A-274	Sequence 274, App
23	2297.5	98.1	2063	12	US-10-017-253A-274	Sequence 274, App
24	2297.5	98.1	2063	12	US-10-173-689-329	Sequence 329, App
25	2297.5	98.1	2063	12	US-10-173-690-329	Sequence 329, App
26	2297.5	98.1	2063	12	US-10-173-691-329	Sequence 329, App
27	2297.5	98.1	2063	12	US-10-173-692-329	Sequence 329, App
28	2297.5	98.1	2063	12	US-10-173-694-329	Sequence 329, App
29	2297.5	98.1	2063	12	US-10-173-698-329	Sequence 329, App
30	2297.5	98.1	2063	12	US-10-173-699-329	Sequence 329, App
31	2297.5	98.1	2063	12	US-10-173-707-329	Sequence 329, App
32	2297.5	98.1	2063	12	US-10-174-569-329	Sequence 329, App
33	2297.5	98.1	2063	12	US-10-174-583-329	Sequence 329, App
34	2297.5	98.1	2063	12	US-10-174-587-329	Sequence 329, App
35	2297.5	98.1	2063	12	US-10-174-589-329	Sequence 329, App
36	2297.5	98.1	2063	12	US-10-174-591-329	Sequence 329, App
37	2297.5	98.1	2063	12	US-10-175-736-329	Sequence 329, App
38	2297.5	98.1	2063	12	US-10-175-743-329	Sequence 329, App
39	2297.5	98.1	2063	12	US-10-175-744-329	Sequence 329, App
40	2297.5	98.1	2063	12	US-10-175-745-329	Sequence 329, App
41	2297.5	98.1	2063	12	US-10-175-748-329	Sequence 329, App
42	2297.5	98.1	2063	12	US-10-175-751-329	Sequence 329, App
43	2297.5	98.1	2063	12	US-10-175-754-329	Sequence 329, App
44	2297.5	98.1	2063	12	US-10-176-480-329	Sequence 329, App
45	2297.5	98.1	2063	12	US-10-176-489-329	Sequence 329, App

ALIGNMENTS

RESULT 1
US-10-180-719-18
; Sequence 18, Application US/10180719
; Publication No. US20030166246A1
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; Hillman, Jennifer L.
; Yue, Henry
; Guegler, Karl J.
; Corley, Neil C.
; Tang, Tom Y.
; Shah, Purvi
; TITLE OF INVENTION: HUMAN PROTEASE MOLECULES
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Dr.
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS

SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/180,719

FILING DATE: 25-Jun-2002

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/09/008,271

FILING DATE: 16-Jan-1998

ATTORNEY/AGENT INFORMATION:

NAME: Mohan-Peterson, Sheila

REGISTRATION NUMBER: 41,201

REFERENCE/DOCKET NUMBER: PF-0458 US

TELECOMMUNICATION INFORMATION:

TELEPHONE: 650-855-0555

TELEFAX: 650-845-4166

INFORMATION FOR SEQ ID NO: 18:

SEQUENCE CHARACTERISTICS:

LENGTH: 2038 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

IMMEDIATE SOURCE:

LIBRARY: COLNOT13

CLONE: 1337018

SEQUENCE DESCRIPTION: SEQ ID NO: 18:

S-10-180-719-18

Lignment Scores:

red. No.:	6,19e-283	Length:	2038
core:	2338, 00	Matches:	434
percent Similarity:	100.00%	Conservative:	1
est Local Similarity:	99.77%	Mismatches:	0
very Match:	99.83%	Indels:	0
B:	12	Gaps:	0

S-09-607-745-2 (1-435) x US-10-180-719-18 (1-2038)

```

Y 1 MetAspProAspSerAspGlnProLeuAsnSerLeuAspValLysProLeuArgLysPro 20
b 200 ATGGATCTCGACGTGATCAACCTCGAAGAGCTCGATGTCAACCCCTCGGCAACCC 259
Y 21 ArgIleProMetGluThrPheArgLysValGlyLeuProIleIleAlaLeuLeuSer 40
b 260 CGTATCCCATGGAGACCTTCAGAAAGTGGGGATCCCATCATCATCATCATCATCTGAGC 319
Y 41 LeuAlaSerIleIleIleValValValLeuLeuLysValIleLeuAspLysTyrPhe 60
b 320 CTGGCGAGTATCATATGTGGTGTCTCTCATCAAGGTGATCTGGATAAATACTACTTTC 379
Y 61 LeuCysGlyGlnProLeuHisPheIleProArgLysGlnLeuCysAspGlyGluLeuAsp 80
b 380 CTCTCGGGGCGACCTCTCCACTTCATCCCGAGGAAGCAGCTGTGTGACGGAGCTGGAC 439
Y 81 CysProLeuGlyGluAspGluHisCysValLysSerPheProGluGlyProAlaVal 100
b 440 TGTCCCTTGGGGGAGGAGGAGGACGTGTGCAAGAGCTTCCCGAAGGGCTCGAGTG 499
Y 101 AlaValArgLeuSerLysAspArgSerThrLeuGlnValLeuAspSerAlaThrGlyAsn 120
b 500 GCAGTCCGCTCTCCAGGACCGATCCACACTGCGAGGTGCTGGACTCGGCCACAGGAA 559
Y 121 TrpPheSerAlaCysPheAspAsnPheThrGluAlaLeuAlaGluThrAlaCysArgGln 140
b 560 TGGTTCTCTGCTGTGTTGACAACTTCAGAAAGCTCTGCTGGAGACAGCTGTAGGCGAG 619
Y 141 MetGlyTyrSerSerLysProThrPheArgAlaValGluIleGlyProAspGlnAspLeu 160
b 620 ATGGGCTACAGCAGCAAAACCCCTTCAGAGCTGGAGATTGGCCACAGCAGATCTG 679
Y 161 AspValValGluIleThrGluAsnSerGlnGluLeuArgMetArgAsnSerSerGlyPro 180
b 680 GATGTTGTTGAATACAGAAACAGCCAGGAGCTTCGATCGGAACTCAAGTGGGCC 739
Y 181 CysLeuSerGlySerLeuValSerLeuHisCysLeuAlaCysGlyLysSerLeuLysThr 200

```

```

Db 740 TGTCTCTCAGGCTCCCTGGTCTCCCTGCACCTGCTTGGGGAGAGCCCTGAAGACC 799
Qy 201 ProArgValValGlyGlyGluGluAlaSerValAspSerTrpProTrpGlnValSerIle 220
Db 800 CCCCGTGTGGTGGTGGGAGGAGGCTCTGTGGATTCTTGGCTTGGCAGGTCAGCATC 859
Qy 221 GlnTyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThr 240
Db 860 CAGTACGACAAACAGCACAGCTCTGTGGAGGAGCATCTCGACCCCACTGGTCTCTCAG 919
Qy 241 AlaAlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySer 260
Db 920 GCAGCCCATCTGCTTCAGGAAACATACCGATGTCTCAACTGGAGGTGCGGAGGCTCA 979
Qy 261 AspLysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleGluPheAsn 280
Db 980 GACAACTGGGAGCTTCCCATCCCTGGCTGTGGCCAGATCATCATTAATTCAAC 1039
Qy 281 ProMetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPhe 300
Db 1040 CCCATGTACCCCAAGACATGACATCGCCCTCATGAAGCTGCAGTTCCCACTCACTTTC 1099
Qy 301 SerGlyThrValArgProIleCysLeuProPheAspGluGluLeuThrProAlaThr 320
Db 1100 TCAGGCACAGTCAGGCCCATCTGTGCCCTTCTTGTAGAGGAGCTCACTCCAGCCACC 1159
Qy 321 ProLeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyGlyLysMetSerAspIle 340
Db 1160 CCACCTCGATCATTTGATGGGCTTTACGACAGATGAGGGAAGATGCTGACATA 1219
Qy 341 LeuLeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAspAlaTyr 360
Db 1220 CTGCTGAGGCGCTCAGTCCAGGTCAITGACAGACACAGCGTCAATGCAGACGATCGCTAC 1279
Qy 361 GlnGlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThr 380
Db 1280 CAGGGGAAGTACCCGAGAAGATGATGTGCGAGGCATCCCGAAGGGGTGTGGACACC 1339
Qy 381 CysGlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGly 400
Db 1340 TGCAGGCTGACAGTGTGGGCCCTGATGTACCAATCTGACAGTGGCATGTGTGGGC 1399
Qy 401 IleValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyValTyrThrLysVal 420
Db 1400 ATCGTTAGCTGGGCTATGGCTGGGGGCCCCAGACCCCGAGGATATACCAAGATC 1459
Qy 421 SerAlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu 435
Db 1460 TCAGCCTATCTCAACTGGATCTACATGTCTGGAAGGCTGAGCTG 1504

```

RESULT 2

US-09-851-588-7

; Sequence 7, Application US/09851588

; Patent No. US20020042067A1

; GENERAL INFORMATION:

; APPLICANT: Mack, David

; APPLICANT: Gish, Kurt C.

; APPLICANT: Wilson, Keith E.

; TITLE OF INVENTION: NOVEL METHODS OF DIAGNOSING COLORECTAL CANCER, COMPOSITIONS, AND

; TITLE OF INVENTION: OF SCREENING FOR COLORECTAL CANCER MODULATORS

; FILE REFERENCE: A-68829-1/DUE/JUD/AMS

; CURRENT APPLICATION NUMBER: US/09/851,588

; CURRENT FILING DATE: 2001-09-24

; PRIOR APPLICATION NUMBER: US 09/642,252

; PRIOR FILING DATE: 2000-08-17

; PRIOR APPLICATION NUMBER: US 09/656,002

; PRIOR FILING DATE: 2000-09-06

; NUMBER OF SEQ ID NOS: 9

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 7

; LENGTH: 2081

; TYPE: DNA

ORGANISM: Homo sapiens
 FEATURE:
 NAME/KEY: CDS
 LOCATION: (215)..(1528)
 OTHER INFORMATION:
 S-09-851-588-7

Alignment Scores:

red. No.: 8.54e-283 Length: 2081
 core: 2337.00 Matches: 434
 Percent Similarity: 100.00% Conservative: 0
 Mismatches: 0
 Indels: 0
 Gaps: 0
 B: 9

S-09-607-745-2 (1-435) x US-09-851-588-7 (1-2081)

```

Y 2 AspProAspSerAspGlnProLeuAsnSerLeuAspVallyysProLeuArgLysProArg 21
O 224 GATCCTGACAGTGATCAACCTCTGAACAGCCTCGATGTCAAACCCCTGCGCAACCCCGT 283
Y 22 IleProMetGlnThrPheArgLysValGlyLeuProIleIleIleAlaLeuLeuSerLeu 41
O 284 ATCCCATGAGAGACCTTCAGAAAGTGGGATCCCATCATCATAGCACTACTGAGCTG 343
Y 42 AlaSerIleIleIleValValLeuIleLysValIleLeuAspLysTyrrPheLeu 61
O 344 GCGAGTATCATTTGTTGTTCTCATCAAGGTGATTCCTGGATAAATACTACTTCTC 403
Y 62 CysGlyGlnProLeuHisPheIleProArgLysGlnLeuCysAspGlyGluLeuAspCys 81
O 404 TGGCGGAGAGCTCTCCACTTCATCCCGAGGAGAGCTGTGTGACGAGAGCTGGATGT 463
Y 82 ProLeuGlyGluAspGluGluHisCysVallyysSerPheProGluGlyProAlaValAla 101
O 464 CCCITGGGGAGGAGGAGGACATCTGTCAAGAGTCTCCCGAGGGCTGCGAGTGCA 523
Y 102 ValArgLeuSerLysAspArgSerThrLeuGlnValLeuAspSerAlaThrGlyAsnTrp 121
O 524 GTCCGCTCTCCAAGGACCGATCCACACTGCGAGTGTGGACTCGGCCACAGGAACTGG 583
Y 122 PheSerAlaCysPheAsnAspPheThrGluAlaLeuAlaGluThrAlaCysArgGlnMet 141
O 584 TTCTCTCGCTGTTTCGACAACTTCACAGAGCTCTCGCTGAGACAGCCTGTAGGACATG 643
Y 142 GlyTyrrSerSerLysProThrPheArgAlaValGluIleGlyProAspGlnAspLeuAsp 161
O 644 GGCTACAGCAGCAACCCACTTTCAGAGCTGTGGAGATTGCCCCAGACAGGATCTGGAT 703
Y 162 ValValGluIleThrGluAsnSerGlnGluLeuArgMetArgAsnSerSerGlyProCys 181
O 704 GTTCTTGAATTCACAGAAAACAGCAGGAGCTTCGATCGCGAACTCAAGTGGGCCCTGT 763
Y 182 LeuSerGlySerLeuValSerLeuHisCysLeuAlaCysGlyLysSerLeuLysThrPro 201
O 764 CTCACAGGCTCCCTGGTCTCCCTGCACTGTCTGCTGTGGAGAGGCTGAGAACCCCC 823
Y 202 ArgValValGlyGlyGluAlaSerValAspSerTrpProTrpGlnValSerIleGln 221
O 824 CGTGTGGTGGGAGGAGGAGGCTCTGTGGATCTTGGCCCTTGGCAGGTGAGCATCCAG 883
Y 222 TyrAspLysGlnHisValCysGlySerIleLeuAspProHisTrpValLeuThrAla 241
O 884 TAGGCAAAACAGCAGCTCTGTGGAGGAGCATCTCTGGACCCCACTGGGTCTCTACGGCA 943
Y 242 AlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaCysSerAsp 261
O 944 GCCCACTGCTTCAGAAACATACCGATGTGTTCAACTGGAAGGTGCGGCGAGCTCAGAC 1003
Y 262 LysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleIleGluPheAsnPro 281
O 1004 AAACCTGGGAGCTTCCCATCCCTGGCTGTGGCCAGATCATCATCATTAATTCAACCCC 1063

```

```

QY 282 MetTyrrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPheSer 301
DB 1064 ATGTACCCCAAGACAAATGACATGCGCCCTCATGAAGCTGCAGTCCACATCTTCTTCA 1123
QY 302 GlyThrValArgProIleCysLeuProPhePheAspGluGluLeuThrProAlaThrPro 321
DB 1124 GGCACAGTCAAGCCCATCTGTCTGCCCTTCTTTGATGAGGAGCTCACTCCAGCCACCCA 1183
QY 322 LeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyGlyLysMetSerAspIleLeu 341
DB 1184 CTCTGGATCATTTGATGGGCTTTACGAAGCAGATGAGGAGGAGATGTCTGACATCTG 1243
QY 342 LeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAspAlaTyrrGln 361
DB 1244 CTGACGCGCTCAGTCAGGTCAATTGACAGCACGCTGCAATGCAGACGATGCTACCAG 1303
QY 362 GlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThrCys 381
DB 1304 GGGAGATCACCGAGAGATGATGTGTGAGGATCTCCGGAAGGGGTGTGGACACCTGC 1363
QY 382 GlnGlyAspSerGlyGlyProLeuMetTyrrGlnSerAspGlnTrpHisValValGlyIle 401
DB 1364 CAGGCTGACAGTGTGGGCCCTGATGTACCAATCTGACAGTGGCATGTGTGGGATC 1423
QY 402 ValSerTrpGlyTyrrGlyCysGlyGlyProSerThrProGlyValTyrrLysValSer 421
DB 1424 GTTAGCTGGGGCTATGGCTGCGGGGCGGAGCACCCAGAGGTATACCAAGGTCTCA 1483
QY 422 AlaTyrrLeuAsnTrpIleTyrrAsnValTrpLysAlaGluLeu 435
DB 1484 GCCTATCTCACTGGATCTACATGTCTGGAAGGTGAGCTG 1525

```

RESULT 3

US-10-097-340-317
 ; Sequence 317, Application US/10097340
 ; Publication No. US20030087250A1
 ; GENERAL INFORMATION:
 ; APPLICANT: JOHN MONAHAN
 ; APPLICANT: Manjula GANNAVAPU
 ; APPLICANT: Sebastian HOERSCH
 ; APPLICANT: Shubhangi KAMATKAR
 ; APPLICANT: Steve G. KOVATS
 ; APPLICANT: Rachel E. MEYERS
 ; APPLICANT: Michael MORRISSEY
 ; APPLICANT: Peter OLANDT
 ; APPLICANT: Ami SEN
 ; APPLICANT: Peter VEIBY
 ; APPLICANT: Gordon B. MILLS
 ; APPLICANT: Robert C. BAST, Jr.
 ; APPLICANT: Karen LU
 ; APPLICANT: Rosemarie SCHMANDT
 ; APPLICANT: Xumei ZHAO
 ; APPLICANT: Karen GLATT
 ; TITLE OF INVENTION: Nucleic Acid Molecules and Proteins For The Identification,
 ; TITLE OF INVENTION: Assessment, Prevention, and Therapy of Ovarian Cancer
 ; FILE REFERENCE: MRI-030
 ; CURRENT APPLICATION NUMBER: US/10/097,340
 ; CURRENT FILING DATE: 2002-03-14
 ; PRIOR APPLICATION NUMBER: 60/276,025
 ; PRIOR FILING DATE: 2001-03-14
 ; PRIOR APPLICATION NUMBER: 60/325,149
 ; PRIOR FILING DATE: 2001-09-26
 ; PRIOR APPLICATION NUMBER: 60/276,026
 ; PRIOR FILING DATE: 2001-03-14
 ; PRIOR APPLICATION NUMBER: 60/324,967
 ; PRIOR FILING DATE: 2001/09/26
 ; PRIOR APPLICATION NUMBER: 60/311,732
 ; PRIOR FILING DATE: 2001-08-10
 ; PRIOR APPLICATION NUMBER: 60/325,102
 ; PRIOR FILING DATE: 2001-09-26
 ; PRIOR APPLICATION NUMBER: 60/323,580
 ; PRIOR FILING DATE: 2001-09-19
 ; NUMBER OF SEQ ID NOS: 363

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 317

LENGTH: 2307

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: misc feature

LOCATION: (1)-(2307)

OTHER INFORMATION: n = A, T, C or G

S-10-097-340-317

Alignment Scores:

red. No.:	1e-282	Length:	2307
core:	2337.00	Matches:	434
percent Similarity:	100.00%	Conservative:	0
est Local Similarity:	100.00%	Mismatches:	0
every Match:	99.79%	Indels:	0
B:	14	Gaps:	0

S-09-607-745-2 (1-435) x US-10-097-340-317 (1-2307)

Y	2	AspProAspSerArgGlnProLeuAsnSerLeuAspValLysProLeuArgLysProArg	21
b	284	GATCCTGACAGTGAATCACTCTGAACAGCTCGATGTCAACCCCTGCGCAACCCCGT	343
Y	22	IleProMetGluThrPheArgLysValGlyLeuProIleLeuLeuLeuSerLeu	41
b	344	ATCCCATGAGACCTTCAGAAAGTGGGATCCCATCATCATAGCAGCTGAGCGCTG	403
Y	42	AlaSerIleLeuValValLeuLeuLysValLeuLeuAspLysTyrTyrPheLeu	61
b	404	GCAGTATCATCTGTGTGTCTCATCAAGGTGATCTGGATAAATACTACTTCTC	463
Y	62	CysGlyGlnProLeuHisPheIleProArgLysGlnLeuCysAspGlyGluLeuAspCys	81
b	464	TGCGGGCAGCTCTCCACTTCATCCCGAGGAGAGCTGTGTGACGAGAGCTGAGCTGT	523
Y	82	ProLeuGlyGluAspGluGluHisCysValLysSerPheProGluGlyProAlaValAla	101
b	524	CCCTTGGGGAGGACGAGGACATGTGTCAAGAGCTTCCCGAGGGCTGCAAGTGGCA	583
Y	102	ValArgLeuSerLysAspArgSerThrLeuGlnValLeuAspSerAlaThrGlyAsnTrp	121
b	584	GTCCGCTCTCCAAGGACCGATCCACACTGCAGCTGTGAGCTCGGCCACAGGAACCTG	643
Y	122	PheSerAlaCysPheAspAsnPhenThrLalaLeuAlaGluThrAlaCysArgGlnMet	141
b	644	TTCTCTGCTCTTTTCGACAACTTCACAGAACTCTCGCTGAGACAGCTGTAGGACATG	703
Y	142	GlyTyrSerSerLysProThrPheArgAlaValGluLeuGlyProAspGlnAspLeuAsp	161
b	704	GGCTACAGCAGCAACCCACTTTCAGAGCTGTGAGATGGCCACAGCAGCACTGGAT	763
Y	162	ValValGluLeuThrGluAsnSerGlnGluLeuArgMetArgAsnSerSerGlyProCys	181
b	764	GTTCGTGAATACACAGAAACAGCAGGAGCTTCGATCGCGAACTCAAGTGGGCGCTGT	823
Y	182	LeuSerGlySerLeuValSerLeuHisCysValLeuAlaCysGlyLysSerLeuLysThrPro	201
b	824	CTCTCAGGCTCCCTGGTCTCCCTGCACTGTCTGCTGTGGAGAGAGCTGAAGACCCCC	883
Y	202	ArgValValGlyGlyGluGluAlaSerValAspSerTrpProTrpGlnValSerIleGln	221
b	884	CGTGTGTGTGTGGGAGAGGCGCTCTGTGATCTTGGCTTGGCAGGTGAGCATCCAG	943
Y	222	TyrAspLysGlnHisValCysGlyGlySerLeuAspProHisTrpValLeuThrAla	241
b	944	TACGACAAACAGCAGCTCTGTGGAGGAGCATCTTGGACCCCCCACTGGGCTCTCAGGCA	1003
Y	242	AlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySerAsp	261
b	1004	GCCCACTGCTTTCAGGAAACATACCGATGTGTCACTGGAAGGTGCGGCGAGGCTCAGAC	1063

QY	262	LysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleLeuPheAsnPro	281
Db	1064	AAACTGGGCGAGCTTCCCATCCCTGGCTGGCGCAAGATCATCATCATTAATCAACCCC	1123
QY	282	MetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPhePheProLeuThrPheSer	301
Db	1124	ATGTACCCCAAGACAAATGACATCGCCCTCATGAAGCTGAGTCCCATCTTCTCA	1183
QY	302	GlyThrValArgProIleCysLeuProPhePheAspGluGluLeuThrProAlaThrPro	321
Db	1184	GGCAGACTCAGGCCCATCTGTCTGCCCTTCTTGTAGAGAGCTCACTCAGCCACCCCA	1243
QY	322	LeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyGlyLysMetSerAspIleLeu	341
Db	1244	CTCTGATCATTTGGATGGGCTTTACGAAGCAGATGGAGGGAAGATGTCTGACATCTG	1303
QY	342	LeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAlaTyrGln	361
Db	1304	CTGAGCGCTCAGTCCAGGTCAATGTGACAGCACCGTCAATGCAGACCATGGTACCAG	1363
QY	362	GlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThrCys	381
Db	1364	GGGAGACTCAGCAGAGAGATGATGTGAGGACATCCCGAGGGGTGTGGACACCTGC	1423
QY	382	GlnGlyAspSerGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGlyIle	401
Db	1424	CAGGTGACAGTGGTGGGCGCTGATGTACCAATCTGACAGGTGGCATGTGGTGGGCATC	1483
QY	402	ValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyValTyrThrLysValSer	421
Db	1484	GTAGTGGGCTATGGCTGCGGGGCGGAGACCCCGAGGATATACCAAGGTCTCA	1543
QY	422	AlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu	435
Db	1544	GCTATCTCACTGATCTACATGTCTGAGAGCTGAGCTG	1585

RESULT 4

US-10-171-311-217

Sequence 217, Application US/10171311

Publication No. US20030087270A1

GENERAL INFORMATION:

APPLICANT: Schlegel, Robert

APPLICANT: Chen, Yan

APPLICANT: Zhao, Xumei

APPLICANT: Monahan, John

APPLICANT: Kamatkar, Shubhangi

APPLICANT: Glatt, Karen

APPLICANT: Gannavarapu, Manjula

APPLICANT: Hoerish, Sebastian

TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY OF CERVICAL CANCER

FILE REFERENCE: MRI-035

CURRENT APPLICATION NUMBER: US/10/171,311

CURRENT FILING DATE: 2002-06-12

PRIOR APPLICATION NUMBER: US 60/298,159

PRIOR FILING DATE: 2001-06-13

PRIOR APPLICATION NUMBER: US 60/298,155

PRIOR FILING DATE: 2001-06-13

PRIOR APPLICATION NUMBER: US 60/335,936

PRIOR FILING DATE: 2001-11-14

NUMBER OF SEQ ID NOS: 238

SOFTWARE: Fast-Seq for Windows Version 4.0

SEQ ID NO 217

LENGTH: 2307

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: misc feature

LOCATION: 1691, 1698, 1705, 1708, 1709, 1713, 1717, 1720, 1724, 1728, 1733, 1741, 1746, 1748, 1755, 1770, 1774, 1791, 1802, 1821, 1838, 1856, 1859, 1864, 1908, 1959, 1997, 2012, 2038, 2143

OTHER INFORMATION: n = A, T, C or G

S-10-171-311-217

Alignment Scores:

red. No.: 1e-282 Length: 2307
core: 2337.00 Matches: 434
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 99.79% Indels: 0
B: 14 Gaps: 0

S-09-607-745-2 (1-435) x US-10-171-311-217 (1-2307)

Y 2 AspProAspSerAspGlnProLeuAsnSerLeuAspValLysProLeuArgLysProArg 21
b 284 GATCTGACGTGATCAACCTCTGACAGCCTCGATGTCAACCCCTCGCAACCCCGT 343
Y 22 IleProMetGluThrPheArgLysValGlyLeuProIleIleIleAlaLeuLeuSerLeu 41
b 344 ATCCCATGGAGACCTTCAGAAAGGTGGGATCCCATCATCATATAGCACTAGGACCTG 403
Y 42 AlaSerIleIleIleValValValLeuLeuLeuValIleLeuAspLysTyrPheLeu 61
b 404 GCGAGTATCATCATGTGTGTCTCATCAAGGTGATTCGGATAAATACTACTTCCTC 463
Y 62 CysGlyGlnProLeuHisPheIleProArgLysGlnLeuCysAspGlyGluLeuAspCys 81
b 464 TCGGGGAGCCTCTCCACTTCATCCCGAGGAGAGCTGTGTGACGGAGCTGGACTGT 523
Y 82 ProLeuGlyGluAspGluGluHisCysValLysSerPheProGluGlyProAlaValAla 101
b 524 CCCITGGGGAGGAGGAGGACATGTGTCAAGACTTCCCGAGGCGCTGCAGTGCA 583
Y 102 ValArgLeuSerLysAspArgSerThrLeuGlnValLeuAspSerAlaThrGlyAsnTrp 121
b 584 GTCCGCTCTCCAAAGACCGATCCACACTGCAGGTGTGGACTCGGCCACAGGAACGTG 643
Y 122 PheSerAlaCysPheAspAsnPheThrGluAlaLeuAlaGluThrAlaCysArgGlnMet 141
b 644 TTCTGTGCTGTTTCGACAACTTCACAGACTCTCGCTGAGACAGCTGTAGGCGAGT 703
Y 142 GlyTyrSerSerLysProThrPheArgAlaValGluIleGlyProAspGlnAspLeuAsp 161
b 704 GGCTACAGCACAAACCACTTCAGACTGTGGAGATTGGCCAGACACAGATCTGGAT 763
Y 162 ValValGluIleThrGluAsnSerGlnGluLeuArgMetArgAsnSerSerGlyProCys 181
b 764 GTTCTTGAATACAGAAACACAGCCAGGAGCTTCGCATGCGGAACTCAAGTGGGCGCTGT 823
Y 182 LeuSerGlySerLeuValSerLeuHisCysLeuAlaCysGlyLysSerLeuLysThrPro 201
b 824 CTCTCAGGCTCCCTGGTCTCCCTGCACTGTCTTGCTGTGGAGAGAGCTGAGACCCCC 883
Y 202 ArgValValGlyGlyGluGluAlaSerValAspSerTrpProTrpGlnValSerIleGln 221
b 884 CGTGTGTGGTGGGGAGGAGGCTCTGTGGATTCTTGCCCTTGGCAGGTTCAGATCCAG 943
Y 222 TyrAspLysGlnHisValCysGlySerLeuLeuAspProHisTrpValLeuThrAla 241
b 944 TACACAAACAGCAGCTGTGGAGGAGCATCTCTGGACCCCACTGGGTCTCCAGGCA 1003
Y 242 AlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySerAsp 261
b 1004 GCCCACTGCTTCAGAAACATACCGATGTGTCACTGGAGAGGTGGGGAGGCTCAGAC 1063
Y 262 LysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleGluPheAsnPro 281
b 1064 AAACTGGGACGCTTCCATCCCTGGCTGTGGCCAGATCATCATCATTAATTCAACCCC 1123
Y 282 MetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPheSer 301
b 1124 ATGTACCCCAAGACAATGACATGCGCCCTCATGAAGCTGAGTTCCTCCACTCATTCTCA 1183
Y 302 GlyThrValArgProIleCysLeuProPhePheAspGluGluLeuThrProAlaThrPro 321

Db 1184 GGCACAGTCAGGCCCATCTGTCTGCCCTTCTTTGATGAGGAGCTCACTCCAGCCACCCCA 1243
Qy 322 LeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyGlyLysMetSerAspIleLeu 341
Db 1244 CTCTGGATCATGTGATGGGCTTTACGAAGCAGAAATGGAGGGAAGATGTCTGACATATCTG 1303
Qy 342 LeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAspAlaTyrGln 361
Db 1304 CTGCAGGCGTCAGTCCAGGTCATTGACAGCACACGGTCAATGCAGACGATGCGTACCAG 1363
Qy 362 GlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThrCys 381
Db 1364 GGGAAAGTCCCGAAGATGATGTGTGAGGCATCCCGAAGGGGTGTGGACACCTGC 1423
Qy 382 GlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGlyIle 401
Db 1424 CAGGTGACAGTGTGTGGGCCCTGATGTACCAATCTGACCACTGGCATGTGTGGGCATC 1483
Qy 402 ValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyValTyrThrLysValSer 421
Db 1484 GTTAGCTGGGCTATGGCTGCGGGGCGCGAGCACCCAGAGTATACACCAAGTCTCA 1543
Qy 422 AlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu 435
Db 1544 GCCTATCTCACTGATCTACATGTCTGGAAGGCTGAGCTG 1585

RESULT 5

US-10-101-510-634
; Sequence 634, Application US/10101510
; Publication No. US20030148295A1
; GENERAL INFORMATION:
; APPLICANT: WANG, YIXIN
; TITLE OF INVENTION: EXPRESSION PROFILES AND METHODS OF USE
; FILE REFERENCE: 15117.0012
; CURRENT APPLICATION NUMBER: US/10101.510
; CURRENT FILING DATE: 2002-03-20
; PRIOR APPLICATION NUMBER: 60/276,947
; PRIOR FILING DATE: 2001-03-20
; NUMBER OF SEQ ID NOS: 805
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 634
; LENGTH: 2165
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-101-510-634

Alignment Scores:
Pred. No.: 9.2e-282 Length: 2165
Score: 2329.00 Matches: 435
Percent Similarity: 99.77% Conservative: 0
Best Local Similarity: 99.77% Mismatches: 0
Query Match: 99.44% Indels: 1
DB: 12 Gaps: 0

US-09-607-745-2 (1-435) x US-10-101-510-634 (1-2165)

Qy 1 MetAspProAspSerAspGlnProLeuAsnSerLeuAspValLysProLeuArgLysPro 20
Db 310 ATGGATCTCTGACAGTGTATCAACCTCTGAAACAGCTCGATGTCAACCCCTGGCGAAACCC 369
Qy 21 ArgIleProMetGluThrPheArgLysValGlyIleProIleIleIleAlaLeuLeuSer 40
Db 370 CGATATCCCATGGAGACCTTCAGAAAGTGGGATCCCATCATCATAGCACTACTGAGC 429
Qy 41 LeuAlaSerIleIleIleValValLeuLeuLysValIleLeuAspLysTyrTyrPhe 60
Db 430 CTGCGAGTATCATCATGTGTGTGTCTCATCAAGGTGATTCGGATAAATACTACTTTC 489
Qy 61 LeuCysGlyGlnProLeuHisPheIleProArgLysGlnLeuCysAspGlyGluLeuAsp 80
Db 490 CTCTGCGGAGGCTCTCCACTTCATCCCGAGGAGCAGCTGTGTGACGGAGAGCTGGAC 549

```
Y      81  CysProLeuGlyGluAspGluHisCysValLysSerPheProGluGlyProAlaVal 100
b      550  TGTCCCTTTGGGGAGAGACGAGAGGACACTGTGTTCAGAGCTTCCCGAAGGGCTCGAGTG 609
Y      101  AlaValArgLeuSerLysAspArgSerThrLeuGlnValLeuAspSerAlaThrGlyAsn 120
b      610  GCAGTCCGCTCTCCAAAGGACCATCCACTGCAGGTGCTGCAGTCCGCCACAGGGAAC 669
Y      121  TrpPheSerAlaCysPheAspAsnSerPheThrGluAlaLeuAlaGluThrAlaCysArgGln 140
b      670  TGGTTCTCTGCTGTTTCACAACTTCACAGAGCTCTCGTGAAGACCTGTGTAGGAG 729
Y      141  MetGlyTyrSerSerLysProThrPheArgAlaValGluLeuLeuLeuLeuLeuLeuLeu 160
b      730  ATGGCTACAGACAGACAAACCCACTTTCAGAGCTGTGGAGATTGGCCACAGCATGTG 789
Y      161  AspValValGluLeuThrGluAsnSerGlnGlu-LeuArgMetArgAsnSerSerGlyPr 180
b      790  GATGTTGTTGAATCACAGAAAACAGCCAGGAGGCTTCGCATGCGGAACCTCAAGTGGCC 849
Y      180  cCysLeuSerGlySerLeuValSerLeuHisCysLeuAlaCysGlyLysSerLeuLysTh 200
b      850  CTGTCTCTCAGGCTCCCTGGTCTCCCTGCACCTGTCTGCTGTGGAGAGCTCGAGAC 909
Y      200  rProArgValValGlyGlyGluGluAlaSerValAspSerTrpProTrpGlnValSeril 220
b      910  CCCCCTGTGTGGTGGGAGAGGCTCTGTGATCTTGGCCCTTGGCAGTCTAGCAT 969
Y      220  eGlnTyrAspLysGlnHisValCysGlyGlySerLeuLeuAspProHisTrpValLeuTh 240
b      970  CCAGTACGACAAACAGACAGCTGTGTGGAGGAGCATCTCTGGACCCCTCACTGGTCTCTAC 1029
Y      240  rAlaAlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValAlaClySe 260
b      1030  GGCAGCCCACTCTCTCAGAAACATACACCATGTGTTCAACTCGAAGGTGCGGGCAGGCTC 1089
Y      260  rAspLysLeuGlySerPheProSerLeuAlaValAlaLysLeuLeuLeuLeuLeuLeu 280
b      1090  AGACAAACTGGGAGCTTCCCATCCCTGCTGTGGCCAGATCATCATTTGATTCAA 1149
Y      280  nProMetTyrProLysAspAsnAspLeuAlaLeuMetLysLeuGlnPheProLeuThrPh 300
b      1150  CCCCATGTACCCCAAGACAAATGACATGCCCTCATGAAGCTGCAGTTCCTCACTT 1209
Y      300  eSerGlyThrValArgProLysCysLeuProPhePheAspGluGluLeuThrProAlaTh 320
b      1210  CTCAGGCACAGTCAGGCCCATCTGTCTGCCCTTCTTTGATGAGGAGCTCACTCCAGCCAC 1269
Y      320  rProLeuTrpLysLeuGlyTrpGlyPheThrLysGlnAsnGlyGlyLysMetSerAspIl 340
b      1270  CCCACTCTGGATCATTTGGATGGGCTTTACGAAGCAGAAATGAGGGAAGATGCTGACAT 1329
Y      340  eLeuLeuGlnAlaSerValGlnValLeuAspSerThrArgCysAsnAlaAspAlaTy 360
b      1330  ACTGTGTGAGGGCTCAGTCCAGTCAATTCACAGCACACGGTGCATTCAGAGAGTATAC 1389
Y      360  rGlnGlyGluValThrGluLysMetMetCysAlaGlyLeuProGluGlyGlyValAspTh 380
b      1390  CCAGGGGGAAGTACCCGGAAGATGATGTGTGAGGACATCCCGAAGGGGGTGTGGACAC 1449
Y      380  rCysGlnGlyAspSerGlyLysProLeuMetTyrGlnSerAspGlnTrpHisValValG 400
b      1450  CTGCCAGGTGACAGTGGTGGGCTCTGATGATACCAATCTGACCATGGCATGTGTGGG 1509
Y      400  yLeValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyValTyrThrLysVa 420
b      1510  CATCGTTAGTGGGCTATGGCTGCGGGGGCCGAGCACCCAGGAGTATACCAAGGT 1569
Y      420  lSerAlaTyrLeuAsnTrpLysValTrpLysAlaGluLeu 435
b      1570  CTCAGCTATCTCACTGATCTACAATGTCTGGAAGGCTGAGCTG 1615
```

```
RESULT 6
US-09-851-588-5
; Sequence 5, Application US/09851588
; Patent No. US20020042067A1
; GENERAL INFORMATION:
; APPLICANT: Mack, David
; APPLICANT: Gish, Kurt C.
; APPLICANT: Wilson, Keith E.
; TITLE OF INVENTION: NOVEL METHODS OF DIAGNOSING COLORECTAL CANCER, COMPOSITIONS, AND P
; TITLE OF INVENTION: OF SCREENING FOR COLORECTAL CANCER MODULATORS
; FILE REFERENCE: A-68829-1/DJE/JUD/AMS
; CURRENT APPLICATION NUMBER: US/09/851,588
; PRIOR FILING DATE: 2001-09-24
; PRIOR APPLICATION NUMBER: US 09/642,252
; PRIOR FILING DATE: 2000-08-17
; PRIOR APPLICATION NUMBER: US 09/656,002
; PRIOR FILING DATE: 2000-09-06
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 2079
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-851-588-5

Alignment Scores:
Pred. No.: 3,67e-281      Length: 2079
Score: 2324.00           Matches: 434
Percent Similarity: 99.77%      Conservative: 0
Best local Similarity: 99.77%      Mismatches: 0
Query Match: 99.23%           Indels: 1
DB: 9                      Gaps: 0
```

US-09-607-745-2 (1-435) x US-09-851-588-5 (1-2079)

```
QY      2  AspProAspSerAspGlnProLeuAsnSerLeuAspValLysProLeuArgLysProArg 21
Db      217  GATCCTGACAGTATCAACCTCTGAACAGCTCGATGTCAAAACCCCTCGCAAAACCCCT 276
QY      22  IleProMetGluThrPheArgLys-ValGlyLeuProIleIleIleAlaLeuSerLe 41
Db      277  ATCCCCATGGAGACCTTCAGAAAGTGTGGGATCCCCATCATCATCATTAATAACTACTTCT 336
QY      41  uAlaSerIleIleIleValValValLeuIleLysValIleLeuAspLysTyrTyrPheLe 61
Db      337  GCGAGTATCATCATTTGGTTGCTCTCATCAAGGTGATTTGGTAAATAACTACTTCTCT 396
QY      61  uCysGlyGlnProLeuHisPheIleProArgLysGlnLeuCysAspGlyGluLeuAspCy 81
Db      397  CTGCGGGCAGCCTCTCCACTTCATCCCGAGGAAGCAGCTGTGTACGAGAGAGCTGGACTG 456
QY      81  sProLeuGlyGluAspGluHisCysValLysSerPheProGluGlyProAlaValAl 101
Db      457  TCCCTTGGGGAGGACGAGGAGGACTGTGTCAAGAGCTTCCCGAAGGECCTGCACTGGC 516
QY      101  aValArgLeuSerLysAspArgSerThrLeuGlnValLeuAspSerAlaThrGlyAsnTr 121
Db      517  AGTCGCGCTCTCCAGGACCCGATCCACACTGCAGGTGCTGGACTCGGCCACAGGGAAC 576
QY      121  pPheSerAlaCysPheAspAsnPheThrGluAlaLeuAlaGluThrAlaCysArgGlnMe 141
Db      577  GTTCTCTCCCTGTTTCGACAACTTCACAGAGCTCTCGCTGAGACAGCTGTAGGACAG 636
QY      141  tGlyTyrSerSerLysProThrPheArgAlaValGluLeuIleGlyProAspGlnAspLeu 161
Db      637  GGGCTACAGACAGAAACCCACTTTCAGAGCTGTGGAGATTGGCCAGACCAAGATCTGGA 696
QY      161  pValValGluIleThrGluAsnSerGlnGluLeuArgMetArgAsnSerSerGlyProCy 181
Db      697  TGTGTTGAATATCAGAAAACAGCCAGGAGCTTCGATGCGGAACTCAAGTGGGCCCTG 756
QY      181  sLeuSerGlySerLeuValSerLeuHisCysLeuAlaCysGlyLysSerLeuLysThrPr 201
```

```
b 757 TCTCTCAGGCTCCCTGGTCTCCCTGCAGCTGCTTGCCTGTGGGAAGAGCCTGAAGACCCC 816
y 201 oArgValValGlyGlyGluGluAlaSerValAspSerTrpProTrpGlnValSerIleGl 221
b 817 CCGTGTGTGGTGGGAGGAGCCCTCTGTGGATCTTGGCCCTTGGCAGGTGAGCATCCA 876
y 221 nTrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThrAl 241
b 877 GTACACAAACAGCACGCTGTGGAGGGAGCATCTGGACCCCTGGTCTCAACGCC 936
y 241 aAlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySerAs 261
b 937 AGCCACATGCTTCAGAAACATACCATGATGTTCACATGGAAGGTGGGGCAGGTTCAGA 996
y 261 pLysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleIleGluPheAsnPr 281
b 997 CAAATGGGACGCTTCCCATCCCTGGCTGTGGCCAAAGATCATCATTAATGAATCAACCC 1056
y 281 oMetTrpProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPheSe 301
b 1057 CATGTACCCCAAGAACATGATCGCCCTCATGAAGCTGCAGTTCGCCACTCACTTTCTTC 1116
y 301 rGlyThrValArgProIleCysLeuProPhePheAspGluGluLeuThrProAlaThrPr 321
b 1117 AGGCACAGTCAGGCCCACTGTCTGCCCTTCTTTGATGAGGAGCTCACTCCAGCCACCCC 1176
y 321 oLeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyGlyLysMetSerAspIleLe 341
b 1177 ACTCTGGATCATTTGGATGGGCTTTACGAAGCAGAAATGGAGGGAAGATGCTTCGACATACT 1236
y 341 tLeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAlaTyrl 361
b 1237 GCTGAGGCGTCAGTCCAGGTCAATGACAGCACAGGTCAATGACAGCATGCGTACCA 1296
y 361 nGlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThrCy 381
b 1297 GGGGGAAGTCACCGAGAAGATGATGTGTCAGGCATCCCGGAGGGGTGTGGACACCTG 1356
y 381 sGlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGlyI 401
b 1357 CCAGGCTGACAGTGTGGGCCCTCTGATGTACCAATCTGACCACTGGCATGTGTGGCAT 1416
y 401 eValSerTrpGlyTrpGlyCysGlyGlyProSerThrProGlyValTrpThrLysValSe 421
b 1417 CATTAGCTGGGCTATGGCTGGGGGGCCGACACCCAGAGATATACCAAGGTCTC 1476
y 421 rAlaTrpLeuAsnTrpIleTyraSnValTrpLysAlaGluLeu 435
b 1477 AGCCTATCTCACTGGATCTCAATGCTTGGAGGCTGAGCTG 1519

RESULT 7
S-09-776-191-71
Sequence 71, Application US/09776191
Publication No. US20030119168A1
GENERAL INFORMATION:
APPLICANT: Edwin L. Madison
APPLICANT: Edgar O. Ong
APPLICANT: Jiunn-Chern Yeh
APPLICANT: Corvas International, Inc.
TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING
TITLE OF INVENTION: TRANSMEMBRANE SERINE PROTEASES, THE ENCODED PROTEINS AND
FILE REFERENCE: 24745-1607
CURRENT APPLICATION NUMBER: US/09/776,191
CURRENT FILING DATE: 2001-02-02
PRIOR APPLICATION NUMBER: 60/213,124
PRIOR FILING DATE: 2000-06-22
PRIOR APPLICATION NUMBER: 60/234,840
PRIOR FILING DATE: 2000-06-22
PRIOR APPLICATION NUMBER: 60/179,982
PRIOR FILING DATE: 2000-02-03
PRIOR APPLICATION NUMBER: 60/183,542
PRIOR FILING DATE: 2000-02-18
```

```

; PRIOR APPLICATION NUMBER: 09/657,968
; PRIOR FILING DATE: 2000-02-08
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 71
; LENGTH: 2079
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (251)...(1522)
; OTHER INFORMATION: Nucleotide sequence encoding transmembrane
; OTHER INFORMATION: protease, serine 4 (TMPRSS4)
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: GenBank NM016425
; DATABASE ENTRY DATE: 2000-11-06
US-09-776-191-71

Alignment Scores:
Pred. No.: 3,67e-281 Length: 2079
Score: 2324.00 Matches: 434
Percent Similarity: 99.77% Conservative: 0
Best Local Similarity: 99.77% Mismatches: 0
Query Match: 99.23% Indels: 1
DB: 11 Gaps: 0

US-09-607-745-2 (1-435) x US-09-776-191-71 (1-2079)
QY 2 AspProAspSerAspGlnProLeuAsnSerLeuAspValLysProLeuArgLysProArg 21
DB 217 GATCTCTGACGTGATCACTCTTGAACAGCTCGATGTCAACCCCTCGGCAACCCCGT 276
QY 22 IleProMetGluThrPheArgLys-ValGlyIleProIleIleIleAlaLeuLeuSerie 41
DB 277 ATCCCATGAGAGACCTTCAGAAAGTGTGGGATCCCATCATCATAGCAGCTACTGAGCCT 336
QY 41 uAlaSerIleIleIleValValLeuIleLysValIleLeuAspLysTyrTyrPheLe 61
DB 337 GCGGAGTATCATCATTTGGTTGCTCTCATCAAGGTGATCTGGATAAATACTACTCTCT 396
QY 61 uCysGlyGlnProLeuHisPheIleProArgLysGlnLeuCysAspGlyGlyLeuAspCy 81
DB 397 CTGGGGCAGCTCTCCACTTCATCCCGAGGAGCAGCTGTGACGAGAGCTGGACTG 456
QY 81 sProLeuGlyGluAspGluGluHisCysValLysSerPheProGluGlyProAlaValAl 101
DB 457 TCCCTTGGGGGAGGACGAGGAGCACTGTGTCAAGAGCTTCCCGAAGGCCCTGCACTGC 516
QY 101 aValArgLeuSerLysAspArgSerThrLeuGlnValLeuAspSerAlaThrGlyAsnTr 121
DB 517 AGTCGCTCTCCAGGACCCGATCCACACTGCGAGGTGTGGACTGGCCACAGGGAACCTG 576
QY 121 pPheSerAlaCysPheAspAsnPheThrGluAlaLeuAlaGluThrAlaCysArgGlnMe 141
DB 577 GTTCTCTGCTGTTTCGACAACTTCACAGAGCTCTCGCTGAGACAGCTGTAGCAGCAT 636
QY 141 tGlyTyrSerSerLysProThrPheArgAlaValGluIleGlyProAspGlnAspLeuAs 161
DB 637 GGGCTACAGCAGCAAAACCCACTTTTCAGAGCTGTGGAGATTGGCCCGACAGCAGGATCTGA 696
QY 161 pValValGluIleThrGluAsnSerGlnGluLeuArgMetArgAsnSerSerGlyProCy 181
DB 697 TGTGTTGAAATTCAGAAAAACAGCCAGGAGCTTCGCAATGCGAACTCAAGTGGGCCCTG 756
QY 181 sLeuSerGlySerLeuValSerLeuHisCysLeuAlaCysGlyLysSerLeuLysThrPr 201
DB 757 TCTCTCAGGCTCCCTGGGTCTCCCTGCACTGTCTGCTGTGGGAGAGCCTGAAGACCCC 816
QY 201 oArgValValGlyGlyGluGluAlaSerValAspSerTrpProTrpGlnValSerIleGl 221
DB 817 CCGTGTGTGGTGGGAGGAGGAGGCTCTGTGGATTTCTTGGCTTGGCAGGTGAGCATCCA 876
QY 221 nTrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThrAl 241
```


997 CAACCTGGGAGCTTCCCATCCCTGGCTGTGGCCAGAGATCATCATTAATTCACACC 1056

281 oMetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPheSe 301

1057 CATGTACCCCAAGACATGATCGCCCTCATGAGCTGCAGTCCCATCTACTTCTC 1116

301 xGlyThrValArgProIleCysLeuProPhePheAspGluLeuThrProLathPr 321

1117 AGGCACAGTCAGGCCCATCTGTCTGCCCTTCTTTGATGAGGAGCTCACTCCAGCACCCC 1176

321 oLeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyClyLysMetSerAspIleLe 341

1177 ACTCTGGATCATTTGATGGGGCTTTTACGAGCAGAAATGGAGGAGAGATGCTGCACATACT 1236

341 uLeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAspAlaTyrGl 361

1237 GCTGCAGGGCTCAGTCCAGGTCATTGACGACACACGGTGCATTCGACACGATGGGTACCA 1296

361 nGlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThrCy 381

1297 GGGGGNAAGTCACCGAGAGAAGATGATGTGTGCAGGCATCCCGAAGGGGGTGTGGACACCTG 1356

381 scI nGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGlyI 401

1357 CCAGGGTCACAGTGTGGGGCCCTCATGTATCCATCTACCAGTGGCATGTGGTGGGCAT 1416

401 eValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyValTyrThrLysValSe 421

1417 CGTTAGCTGGGGCTATGGCTGGGGGGGCCGAGCACCCAGAGGTATACACCAAGGTCCTC 1476

421 rAlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu 435

1477 AGCCCTATCTCACTGGATCTCAAAATGTCTCGAAGCTCAGGTG 1519

RESULT 9

3-10-254-289-1
 Sequence 1, Application US/10254289
 Publication NO. US20030118509A1
 GENERAL INFORMATION:
 APPLICANT: Mack, David
 APPLICANT: Gish, Kurt
 APPLICANT: Wilson, Keith
 TITLE OF INVENTION: NOVEL METHODS OF DIAGNOSING COLORECTAL CANCER, COMPOSITIONS, AND
 TITLE OF INVENTION: OF SCREENING FOR COLORECTAL CANCER MODULATORS
 FILE REFERENCE: A-69108/DJBE/JJD/RMS
 CURRENT APPLICATION NUMBER: US/10/254,289
 CURRENT FILING DATE: 2002-09-24
 PRIOR APPLICATION NUMBER: US/09/656,002
 PRIOR FILING DATE: 2000-09-06
 PRIOR APPLICATION NUMBER: US 09/525,993
 PRIOR FILING DATE: 2000-03-15
 PRIOR APPLICATION NUMBER: US 09/493,444
 PRIOR FILING DATE: 2000-01-28
 PRIOR APPLICATION NUMBER: PCT/US 00/07044
 PRIOR FILING DATE: 2000-03-15
 NUMBER OF SEQ ID NOS: 3
 SOFTWARE: PatentIn version 3.0
 SEQ ID NO 1
 LENGTH: 2079
 TYPE: DNA
 ORGANISM: Homo sapiens
 3-10-254-289-1
 Alignment Scores:
 red. No.: 3,67e-281 Length: 2079
 core: 2324.00 Matches: 434
 Percent Similarity: 99.77% Conservative: 0
 Percent Local Similarity: 99.77% Mismatches: 0
 Query Match: 99.23% Indels: 1
 3: 14 Gaps: 0
 3-09-607-745-2 (1-435) x US-10-254-289-1 (1-2079)

361 nGlyGluValThrGluLysMetMetCysAlaGlyLeuProGluGlyValAspThrCy 381
 1297 GGGGAAGTCAACGAGAGATGATGTGTGAGGATCCCGAAGGGGTGTGACACCTG 1356
 381 sGlnGlyAspSerGlyGlyProLeuMetTyGlnSerAspGlnTrpHisValValGlyI 401
 1357 CCAGGGTGCAGTGGTGGGCCCCGTGATGATACCAATCTGACAGTGGCATGTGGTGGGCAT 1416
 401 eValSerTrpGlyTyGlyCysGlyGlyProSerThrProGlyValTyThrLysValse 421
 1417 CGTTCAGTGGGCTATGGCTGGGGGGCCGAGACCCCGAGGATATACACCAAGGTCTC 1476
 421 rAlaTyLeuAsnTrpLeuTyAsnValTrpLysAlaGluLeu 435
 1477 AGCCTATCTCAACTGGATCTACAATGTCTGGAAGGCTGAGCTG 1519

RESULT 10

S-09-776-191-3
 Sequence 3, Application US/09776191
 Publication No. US20030119168A1
 GENERAL INFORMATION:
 APPLICANT: Edwin L. Madison
 APPLICANT: Edgar O. Ong
 APPLICANT: Jiumn-Chern Yeh
 APPLICANT: Corvas International, Inc.
 TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING
 TITLE OF INVENTION: TRANSMEMBRANE SERINE PROTEASES, THE ENCODED PROTEINS AND
 TITLE OF INVENTION: METHODS BASED THEREON
 FILE REFERENCE: 24745-1607
 CURRENT APPLICATION NUMBER: US/09/776,191
 CURRENT FILING DATE: 2001-02-02
 PRIOR APPLICATION NUMBER: 60/213,124
 PRIOR FILING DATE: 2000-06-22
 PRIOR APPLICATION NUMBER: 60/234,840
 PRIOR FILING DATE: 2000-06-22
 PRIOR APPLICATION NUMBER: 60/179,982
 PRIOR FILING DATE: 2000-02-03
 PRIOR APPLICATION NUMBER: 60/183,542
 PRIOR FILING DATE: 2000-02-18
 PRIOR APPLICATION NUMBER: 09/657,968
 PRIOR FILING DATE: 2000-02-08
 NUMBER OF SEQ ID NOS: 72
 SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 3
 LENGTH: 2137
 TYPE: DNA
 ORGANISM: Homo Sapien
 FEATURE:
 NAME/KEY: CDS
 LOCATION: (261)...(1574)
 OTHER INFORMATION: DNA sequence encoding a transmembrane serine
 OTHER INFORMATION: protease (MSP3) protein
 S-09-776-191-3

S-09-776-191-3

alignment Scores:
 red. No.: 1.63e-280 Length: 2137
 core: 2319.00 Matches: 432
 ercent Similarity: 99.54% Conservative: 0
 est Local Similarity: 99.54% Mismatches: 2
 uery Match: 99.02% Indels: 0
 B: 11 Gaps: 0

S-09-607-745-2 (1-435) x US-09-776-191-3 (1-2137)

Y 2 AspProAspSerAspGlnProLeuAsnSerLeuAspValLysProLeuArgLysProArg 21
 b 270 GATCTGTACAGTGAATCAACCTCTGAACAGCCTCGATGTCAACCCCTGCGAAACCCCGT 329
 Y 22 IleProMetGluThrPheArgLysValGlyIleProLleIleIleAlaLeuSerLeu 41
 b 330 ATCCCCATGGAGACCTTCAGAAAGTGGGGATCCCATCATCATAGCACTACTGAGCCTG 389
 Y 42 AlaSerIleIleIleValValValLeuLeuLysValIleLeuAspLysTyTrPheLeu 61

Db 390 GGGAGTATCATATTGTGGTGTCTTCATCAAGGTGATTCTGGATAAAATACTACTTCCTC 449
 QY 62 CysGlyGlnProLeuHisPheIleProArgLysGlnLeuCysAspGlyGluLeuAspCys 81
 Db 450 TGGGGGACGCTCTCCATCTTCATCCGAGGAGACAGCTGTGTGACGAGAGCTGGACTGT 509
 QY 82 ProLeuGlyGluAspGluGluHisCysValLysSerPheProGluGlyProAlaValAla 101
 Db 510 CCCTTGGGGAGGACGAGGAGACACTGTGTCAAGAGCTTCCCCGAAGGGCCCTGAGTGGCA 569
 QY 102 ValArgLeuSerLysAspArgSerThrLeuGlnValLeuAspSerAlaThrGlyAsnTrp 121
 Db 570 GTCCGCTCTCCAAAGACCGATCCACACTGCAGGTGCTGACTCGGCCACAGGGAAGTGG 629
 QY 122 PheSerAlaCysPheAspAsnPheThrGluAlaLeuAlaGluThrAlaCysArgGlnMet 141
 Db 630 TTCTCTGCTGTTTCGACCACTTCACAGAGCTCTCGCTGAGACAGCCTGTAGGCAGATG 689
 QY 142 GlyTyTrSerSerLysProThrPheArgAlaValGluIleGlyProAspGlnAspLeuAsp 161
 Db 690 GGCTACAGCAGCAAAACCCACCTTCAGAGCTGTGGAGATTGGCCCGACAGCAGGATCTGGAT 749
 QY 162 ValValGluIleThrGluAsnSerGlnGluLeuArgMetArgAsnSerSerGlyProCys 181
 Db 750 GTTGTGTAATCACAGAAACAGCCAGGAGCTTCGCATGCGGAACCTCAAGTGGGCCCTGT 809
 QY 182 LeuSerGlySerLeuValSerLeuHisCysLeuAlaCysGlyLysSerLeuLysThrPro 201
 Db 810 CTCTCAGGCTCCCTGCTCTCCCTGCACTGTCTGCTGTGGGAGAGCCTGAAGACCCCC 869
 QY 202 ArgValValGlyGlyGluGluAlaSerValAspSerTrpProTrpGlnValSerIleGln 221
 Db 870 CGTGTGGTGGTGGGAGGAGGCTCTGTGGATTCTTGGCCTTGGCAGGTCAGCATCCAG 929
 QY 222 TyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThrAla 241
 Db 930 TAGCATATACAGCACGCTGTGTGGAGGAGCATCTGGACCCCACTGGGTGCTCACCAGCA 989
 QY 242 AlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySerAsp 261
 Db 990 GCCCAGCTGTTCCAGGAAACATACCGATGTCTCACTGGAGGTGCGGGCAGGCTCAGAC 1049
 QY 262 LysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleLeuGluPheAsnPro 281
 Db 1050 AAATGGGCAGCTTCCATCCCTGGCTGTGGCCAGATCATCATCATTTGAATTCACACCC 1109
 QY 282 MetTyTrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPheSer 301
 Db 1110 ATGTACCCCAAGACAAATGACATCGCCCTCATGAAGCTGAGTTCACCTTCTCTCA 1169
 QY 302 GlyThrValArgProLleCysLeuProPheAspGluGluLeuThrProAlaThrPro 321
 Db 1170 GGCACAGTCAGGCTCATCTCTGCTGCTTCTTTGATGAGGAGCTCACTCCAGCACCCCA 1229
 QY 322 LeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyGlyLysMetSerAspIleLeu 341
 Db 1230 CTCTGGATCATTTGGATGGGCTTTTACGAGCAGAAATGAGGAGAGATGTCTGTGACATATG 1289
 QY 342 LeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAspAlaTyGln 361
 Db 1290 CTCAGGCGTCAGTCCAGGTCAATGTGACAGCACAGGTGCAATGCAGACGATGGGTACCA 1349
 QY 362 GlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThrCys 381
 Db 1350 GGGGAAGTCAACGAGAGATGATGTGTGAGGATCCCGAAGGGGTGTGGACACCTGC 1409
 QY 382 GlnGlyAspSerGlyGlyProLeuMetTyGlnSerAspGlnTrpHisValValGlyIle 401
 Db 1410 CAGGGTCAGAGTGGTGGGCCCTGTGATCAATCTCAGCAGTGGCATGTGTGGGGCATC 1469
 QY 402 ValSerTrpGlyTyGlyCysGlyGlyProSerThrProGlyValTyThrLysValSer 421

382 GlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGlyIle 401
1344 CAGGGTGACAGTGGTGGGCCCTGATGTACCAATCTGACCACTGGCAGTGGTGGGCATC 1403
402 ValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyValTyrThrLysValSer 421
1404 GTTAGCTGGGGCTATGGTGGGGGGCCCGAGCACCCAGAGGTATACACCAAGGTCTCA 1463
422 AlaTyrIeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu 435
1464 GCCTATCTCAACTGATCTACAATGCTCTGGAAGGTGAGCTG 1505

RESULT 12

S-09-946-374-274

Sequence 274, Application US/09946374

Publication No. US20030073129A1

GENERAL INFORMATION:

APPLICANT: Baker, Kevin P.
APPLICANT: Botstein, David
APPLICANT: Desnoyers, Luc
APPLICANT: Baton, Dan L.
APPLICANT: Ferrara, Napoleone
APPLICANT: Fong, Sherman
APPLICANT: Gao, Wei-Qiang
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Grimaldi, Christopher J.
APPLICANT: Gurney, Austin L.
APPLICANT: Hillan, Kenneth J.
APPLICANT: Pan, James
APPLICANT: Paoni, Nicholas F.
APPLICANT: Roy, Margaret Ann
APPLICANT: Smith, Victoria
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Watanabe, Colin K.
APPLICANT: Williams, P. Mickey
APPLICANT: Wood, William I.

TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
Acids Encoding the Same

FILE REFERENCE: P2830P1C1

CURRENT APPLICATION NUMBER: US/09/946,374

CURRENT FILING DATE: 2001-09-04

PRIOR APPLICATION NUMBER: 60/098716

PRIOR FILING DATE: 1998-09-01

PRIOR APPLICATION NUMBER: 60/098723

PRIOR FILING DATE: 1998-09-01

PRIOR APPLICATION NUMBER: 60/098749

PRIOR FILING DATE: 1998-09-01

PRIOR APPLICATION NUMBER: 60/098750

PRIOR FILING DATE: 1998-09-01

PRIOR APPLICATION NUMBER: 60/098803

PRIOR FILING DATE: 1998-09-02

PRIOR APPLICATION NUMBER: 60/098821

PRIOR FILING DATE: 1998-09-02

PRIOR APPLICATION NUMBER: 60/098843

PRIOR FILING DATE: 1998-09-02

PRIOR APPLICATION NUMBER: 60/099536

PRIOR FILING DATE: 1998-09-09

PRIOR APPLICATION NUMBER: 60/099596

PRIOR FILING DATE: 1998-09-09

PRIOR APPLICATION NUMBER: 60/099598

PRIOR FILING DATE: 1998-09-09

PRIOR APPLICATION NUMBER: 60/099602

PRIOR FILING DATE: 1998-09-09

PRIOR APPLICATION NUMBER: 60/099642

PRIOR FILING DATE: 1998-09-09

PRIOR APPLICATION NUMBER: 60/099741

PRIOR FILING DATE: 1998-09-10

PRIOR APPLICATION NUMBER: 60/099754

PRIOR FILING DATE: 1998-09-10

PRIOR APPLICATION NUMBER: 60/099763

PRIOR FILING DATE: 1998-09-10

PRIOR APPLICATION NUMBER: 60/099792
PRIOR FILING DATE: 1998-09-10
PRIOR APPLICATION NUMBER: 60/099808
PRIOR FILING DATE: 1998-09-10
PRIOR APPLICATION NUMBER: 60/099812
PRIOR FILING DATE: 1998-09-10
PRIOR APPLICATION NUMBER: 60/099815
PRIOR FILING DATE: 1998-09-10
PRIOR APPLICATION NUMBER: 60/099816
PRIOR FILING DATE: 1998-09-10
PRIOR APPLICATION NUMBER: 60/100385
PRIOR FILING DATE: 1998-09-15
PRIOR APPLICATION NUMBER: 60/100388
PRIOR FILING DATE: 1998-09-15
PRIOR APPLICATION NUMBER: 60/100390
PRIOR FILING DATE: 1998-09-15
PRIOR APPLICATION NUMBER: 60/100584
PRIOR FILING DATE: 1998-09-16
PRIOR APPLICATION NUMBER: 60/100627
PRIOR FILING DATE: 1998-09-16
PRIOR APPLICATION NUMBER: 60/100661
PRIOR FILING DATE: 1998-09-16
PRIOR APPLICATION NUMBER: 60/100662
PRIOR FILING DATE: 1998-09-16
PRIOR APPLICATION NUMBER: 60/100664
PRIOR FILING DATE: 1998-09-16
PRIOR APPLICATION NUMBER: 60/100683
PRIOR FILING DATE: 1998-09-17
PRIOR APPLICATION NUMBER: 60/100684
PRIOR FILING DATE: 1998-09-17
PRIOR APPLICATION NUMBER: 60/100710
PRIOR FILING DATE: 1998-09-17
PRIOR APPLICATION NUMBER: 60/100711
PRIOR FILING DATE: 1998-09-17
PRIOR APPLICATION NUMBER: 60/100848
PRIOR FILING DATE: 1998-09-18
PRIOR APPLICATION NUMBER: 60/100849
PRIOR FILING DATE: 1998-09-18
PRIOR APPLICATION NUMBER: 60/100919
PRIOR FILING DATE: 1998-09-17
PRIOR APPLICATION NUMBER: 60/100930
PRIOR FILING DATE: 1998-09-17
PRIOR APPLICATION NUMBER: 60/101014
PRIOR FILING DATE: 1998-09-18
PRIOR APPLICATION NUMBER: 60/101068
PRIOR FILING DATE: 1998-09-18
PRIOR APPLICATION NUMBER: 60/101071
PRIOR FILING DATE: 1998-09-18
PRIOR APPLICATION NUMBER: 60/101279
PRIOR FILING DATE: 1998-09-22
PRIOR APPLICATION NUMBER: 60/101471
PRIOR FILING DATE: 1998-09-23
PRIOR APPLICATION NUMBER: 60/101472
PRIOR FILING DATE: 1998-09-23
PRIOR APPLICATION NUMBER: 60/101474
PRIOR FILING DATE: 1998-09-23
PRIOR APPLICATION NUMBER: 60/101475
PRIOR FILING DATE: 1998-09-23
PRIOR APPLICATION NUMBER: 60/101476
PRIOR FILING DATE: 1998-09-23
PRIOR APPLICATION NUMBER: 60/101477
PRIOR FILING DATE: 1998-09-23
PRIOR APPLICATION NUMBER: 60/101479
PRIOR FILING DATE: 1998-09-23
PRIOR APPLICATION NUMBER: 60/101738
PRIOR FILING DATE: 1998-09-24
PRIOR APPLICATION NUMBER: 60/101741
PRIOR FILING DATE: 1998-09-24
PRIOR APPLICATION NUMBER: 60/101743
PRIOR FILING DATE: 1998-09-24
PRIOR APPLICATION NUMBER: 60/101915
PRIOR FILING DATE: 1998-09-24
PRIOR APPLICATION NUMBER: 60/101916

PRIOR FILING DATE: 1998-09-24
PRIOR APPLICATION NUMBER: 60/102207
PRIOR FILING DATE: 1998-09-29
PRIOR APPLICATION NUMBER: 60/102240
PRIOR FILING DATE: 1998-09-29
PRIOR APPLICATION NUMBER: 60/102307
PRIOR FILING DATE: 1998-09-29
PRIOR APPLICATION NUMBER: 60/102330
PRIOR FILING DATE: 1998-09-29
PRIOR APPLICATION NUMBER: 60/102331
PRIOR FILING DATE: 1998-09-29
PRIOR APPLICATION NUMBER: 60/102484
PRIOR FILING DATE: 1998-09-30
PRIOR APPLICATION NUMBER: 60/102487
PRIOR FILING DATE: 1998-09-30
PRIOR APPLICATION NUMBER: 60/102570
PRIOR FILING DATE: 1998-09-30
PRIOR APPLICATION NUMBER: 60/102571
PRIOR FILING DATE: 1998-09-30
PRIOR APPLICATION NUMBER: 60/102684
PRIOR FILING DATE: 1998-10-01
PRIOR APPLICATION NUMBER: 60/102687
PRIOR FILING DATE: 1998-10-01
PRIOR APPLICATION NUMBER: 60/102965
PRIOR FILING DATE: 1998-10-02
PRIOR APPLICATION NUMBER: 60/103258
PRIOR FILING DATE: 1998-10-06
PRIOR APPLICATION NUMBER: 60/103314
PRIOR FILING DATE: 1998-10-07
PRIOR APPLICATION NUMBER: 60/103315
PRIOR FILING DATE: 1998-10-07
PRIOR APPLICATION NUMBER: 60/103328
PRIOR FILING DATE: 1998-10-07
PRIOR APPLICATION NUMBER: 60/103395
PRIOR FILING DATE: 1998-10-07
PRIOR APPLICATION NUMBER: 60/103396
PRIOR FILING DATE: 1998-10-07
PRIOR APPLICATION NUMBER: 60/103401
PRIOR FILING DATE: 1998-10-07
PRIOR APPLICATION NUMBER: 60/103449
PRIOR FILING DATE: 1998-10-06
PRIOR APPLICATION NUMBER: 60/103633
PRIOR FILING DATE: 1998-10-08
PRIOR APPLICATION NUMBER: 60/103678
PRIOR FILING DATE: 1998-10-08
PRIOR APPLICATION NUMBER: 60/103679
PRIOR FILING DATE: 1998-10-08
PRIOR APPLICATION NUMBER: 60/103711
PRIOR FILING DATE: 1998-10-08
PRIOR APPLICATION NUMBER: 60/104257
PRIOR FILING DATE: 1998-10-14
PRIOR APPLICATION NUMBER: 60/104987
PRIOR FILING DATE: 1998-10-20
PRIOR APPLICATION NUMBER: 60/105000
PRIOR FILING DATE: 1998-10-20
PRIOR APPLICATION NUMBER: 60/105002
PRIOR FILING DATE: 1998-10-20
PRIOR APPLICATION NUMBER: 60/105104
PRIOR FILING DATE: 1998-10-21
PRIOR APPLICATION NUMBER: 60/105169
PRIOR FILING DATE: 1998-10-22
PRIOR APPLICATION NUMBER: 60/105266
PRIOR FILING DATE: 1998-10-22
PRIOR APPLICATION NUMBER: 60/105693
PRIOR FILING DATE: 1998-10-25
PRIOR APPLICATION NUMBER: 60/105694
PRIOR FILING DATE: 1998-10-26
PRIOR APPLICATION NUMBER: 60/105807

alignent Scores:

red. No.:	7, 76e-278	Length:	2063
core:	2297.50	Matches:	429
arcent Similarity:	98.85%	Conservative:	0

Best Local Similarity:	98.85%	Mismatches:	0
Query Match:	98.10%	Indels:	5
DB:	11	Gaps:	1
US-09-607-745-2 (1-435) x US-09-946-374-274 (1-2063)			
Qy	2	AspProAspSerAspGlnProLeuAsnSerLeuAspVallysProLeuArgLysProArg	21
Db	219	GATCCTCACAGTATCAACCTCTGAACAGCTCGATGTCAACCCCTCGCAAAACCCCGT	278
Qy	22	IleProMetGluThrPheArgLysValGlyLeuProIleIleLeuAlaLeuSerLeu	41
Db	279	ATCCCATGGAGACCTTCAGAAAGTGGGATCCCATCATCATATAGCACACTACTGAGCTG	338
Qy	42	AlaSerIleIleIleValValLeuIleLysValIleLeuAspLysTyrTyrPheLeu	61
Db	339	GCGAGTATCATATTGGTGTTCCTCATCAAGTGATTCTGGATTAATACTACTTCTCTC	398
Qy	62	CysGlyGlnProLeuHisPheIleProArgLysGlnLeuCysAspGlyGluLeuAspCys	81
Db	399	TGCGGGCAGCTCTCCACTTCATCCGAGGAAGCAGCTGTGTGACGGAGAGCTGACTGT	458
Qy	82	ProLeuGlyGluAspGluGluHisCysValIysSerPheProGluGlyProAlaValAla	101
Db	459	CCCTTGGGGGAGGACGAGGAGCACTGTGTCAAGAGCTTCCCGAAGGGCTTGCAGTGGCA	518
Qy	102	ValArgLeuSerLysAspArgSerThrLeuGlnValLeuAspSerAlaThrGlyAsnTrp	121
Db	519	GTCCGCTCTCTCAAGGACCGATCCACACTGCAGTGTGTGACTCGGCCACAGGAACCTGG	578
Qy	122	PheSerAlaCysPheAspAsnPheThrGluAlaLeuAlaGluThrAlaCysArgGlnMet	141
Db	579	TTCTCTCCCTGTTTCGACAACTTCACAGAAAGCTCTCGCTGAGACAGCTGTAGCAGATG	638
Qy	142	GlyTyrSerSerLysProThrPheArgAlaValGluIleGlyProAspGlnAspLeuAsp	161
Db	639	GGCTACAGC-----AGAGCTGTGGAGATTGGCCGACAGGATCTGGAT	683
Qy	162	ValValGluIleThrGluAsnSerGlnGluLeuArgMetArgAsnSerSerGlyProCys	181
Db	684	GTGTGTGAATCACAGAAACAGCCAGGAGCTTCGCATGCGGAATCAAGTGGGCCCTGT	743
Qy	182	LeuSerGlySerLeuValSerLeuHisCysLeuAlaCysGlyLysSerLeuLysThrPro	201
Db	744	CTCTCAGGCTCCCTGCTCTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	803
Qy	202	ArgValValGlyGlyGluGluAlaSerValAspSerTrpProTrpGlnValSerIleGln	221
Db	804	CGTGTGTGGTGGGAGGAGGCTCTGTGATTTCTTGGCTTGGCAGGTGCGATCCAG	863
Qy	222	TyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThrala	241
Db	864	TACGACAAACAGCACGCTGTGGAGGAGGATCTCGGACCCCTGCTGCTGCTGCTGCTGCT	923
Qy	242	AlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySerAsp	261
Db	924	GCCCACTGCTTCAGAAACATACCGATGTGTTCACCTGGAAGGTGCGGGAGGCTCAGAC	983
Qy	262	LysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleGluPheAsnPro	281
Db	984	AACTGGGAGCTTCCCATCTCCCTGCTGGCCAAAGATCATCATCATTAATTCACCCCT	1043
Qy	282	MetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPheSer	301
Db	1044	ATGTACCCCAAGACATGATCGCCCTCATGAAGCTGCAGTTCCTCCACTCACCTTCTCA	1103
Qy	302	GlyThrValArgProIleCysLeuProPhePheAspGluGluLeuThrProAlaThrPro	321
Db	1104	GGCACAGTCAGGCCCATCTGTCTGCCCTTTTGTATGAGGAGCTCATCTCCAGCCACCCCA	1163
Qy	322	LeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyGlyLysMetSerAspIleLeu	341
Db	1164	CTCTGGATCATTTGGATGGGCTTTTACGAAGCAGAAATGGAGGAAGATGTCTGACATACTG	1223

y 342 LeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAlaTyrGln 361
 b 1224 CTGAGGGCGTCAGTCCAGTCAATTGACAGCACACGGTGCAATGCAGACGATGCGGTACACG 1283
 y 362 GlyGlnValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThrCys 381
 b 1284 GGGGAAGTCACCGAAGATGATGTGTGACGACATCCCGAAGGGGTGTGACACCTGC 1343
 y 382 GlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGlnTyrHisValValGlyIle 401
 b 1344 CAGGTTGACATGTGTGGCCCTCATATGATACCACTGACAGTGGCATGTGTGGGCATC 1403
 y 402 ValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyValTyrThrLysValSer 421
 b 1404 GTTAGCTGGGCTATGGCTGCGGGGCCCGAGACCCCGAGAGTATACACCAAGGTCTCA 1463
 y 422 AlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu 435
 b 1464 GCCTATCTCAACTGATCTACAATGTCTGGAAGGCTGAGCTG 1505

ESULT 13

S-10-015-387A-274

Sequence 274, Application US/10015387A

Publication No. US20030135034A1

GENERAL INFORMATION:

APPLICANT: Baker, Kevin P.
 APPLICANT: Botstein, David
 APPLICANT: Desnoyers, Luc
 APPLICANT: Eaton, Dan I.
 APPLICANT: Ferrara, Napoleone
 APPLICANT: Fong, Sherman
 APPLICANT: Gao, Wei-Qiang
 APPLICANT: Goddard, Audrey
 APPLICANT: Godowski, Paul J.
 APPLICANT: Grimaldi, Christopher J.
 APPLICANT: Gurney, Austin L.
 APPLICANT: Hillan, Kenneth J.
 APPLICANT: Pan, James
 APPLICANT: Paoni, Nicholas F.
 TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
 FILE REFERENCE: P2830PIC54
 CURRENT APPLICATION NUMBER: US/10/015,387A
 CURRENT FILING DATE: 2001-12-12
 Prior Application removed - See File Wrapper or Palm
 NUMBER OF SEQ ID NOS: 477
 SEQ ID NO 274
 LENGTH: 2063
 TYPE: DNA

ORGANISM: Homo sapiens

S-10-015-387A-274

Alignment Scores:
 red. No.: 7,76e-278 Length: 2063
 core: 2297.50 Matches: 429
 percent Similarity: 98.85% Conservative: 0
 est Local Similarity: 98.85% Mismatches: 0
 uery Match: 98.10% Indels: 5
 B: 12 Gaps: 1

S-09-607-745-2 (1-435) x US-10-015-387A-274 (1-2063)

y 2 AspProAspSerAspGlnProLeuAsnSerLeuAspValLysProLeuArgLysProArg 21
 b 219 GATCCTGACAGTATCAACCTCTGAACAGCTCGATGTCAAACCCCTCGGCAACCCCGT 278
 y 22 IleProMetGluThrPheArgLysValGlyIleProIleIleAlaLeuLeuSerLeu 41
 b 279 ATCCCATGGAGACCTTCAGAAAGGTGGGATCCCATCATCATAGCACTACTAGGCTG 338
 y 42 AlaSerIleIleValValValLeuIleLysValIleLeuAspLysTyrPheLeu 61

Db 339 GCGAGTATCATCATTTGTGGTGTCTCTCATCAAGGTGATTTCTGGATAAATACTACTTCTCTC 398
 Qy 62 CysGlyGlnProLeuHisPheIleProArgLysGlnLeuCysAspGlyGluLeuAspCys 81
 Db 399 TGGGGGAGAGCTCTCCACTTCATCCCGAGGAAGACAGCTGTGTGACGGAGAGCTGACTGT 458
 Qy 82 ProLeuGlyGlnAspGluGluHisCysValLysSerPheProGluGlyProAlaValAla 101
 Db 459 CCCTTGGGGGAGACGAGGAGACATGTGTCAAGAGCTTCCCCGAAGGGCCCTGCAGTGGCA 518
 Qy 102 ValArgLeuSerLysAspArgSerThrLeuGlnValLeuAspSerAlaThrGlyAsnTrp 121
 Db 519 GTCCGCTCTCCAGGACCGATCCACTCGAGTGTGTGACTCGGCCACAGGAACTGG 578
 Qy 122 PheSerAlaCysPheAspAsnPheThrGluAlaLeuAlaGluThrAlaCysArgGlnMet 141
 Db 579 TTCTCTGCTGTTTGGCAACTTCACAGAAGCTCTCGCTGAGACAGCTGTGAGGCAGATG 638
 Qy 142 GlyTyrSerSerLysProThrPheArgAlaValGluIleGlyProAspGlnAspLeuAsp 161
 Db 639 GGCTACAGC-----AGAGCTGTGGAGATTGGCCCGACAGGATCTGGAT 683
 Qy 162 ValValGluIleThrGluAsnSerGlnGluLeuArgMetArgAsnSerSerGlyProCys 181
 Db 684 GTTGTGAAATCAGAAACAGCCAGAGCTTGCATGCGAACTCAAGTGGGCCCTGT 743
 Qy 182 LeuSerGlySerLeuValSerLeuHisCysLeuAlaCysGlyLysSerLeuLysThrPro 201
 Db 744 CTCTCAGGCTCCCTGGTCTCCCTGCATCTGTCTGCTGTGGAGAGACCTCAAGACCCCC 803
 Qy 202 ArgValValGlyGlyGluGluAlaSerValAspSerTrpProTrpGlnValSerIleGln 221
 Db 804 CGTGTGTGGGTGGGAGGAGGCTCTGTGTGATTTCTTGGCTTGGCAGTGCAGATCCAG 863
 Qy 222 TyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThrAla 241
 Db 864 TACGACAAACAGACAGCTCTGTGTGAGGAGCATCTCTGGACCCCTCTGGTCTCTCAGGCA 923
 Qy 242 AlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySerAsp 261
 Db 924 GCCCAGCTGCTTCAGAAACATACCGATGTGTCAACTGGAAGGTGGCGGAGGCTCAGAC 983
 Qy 262 LysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleGluPheAsnPro 281
 Db 984 AAACCTGGGCGAGTTCCTCATCCCTGGCTGTGGCCAAAGATCATCATTAATTCACACCC 1043
 Qy 282 MetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPheSer 301
 Db 1044 ATGTACCCCAAGACATGACATCGCCCTCATGNAGCTGCAGTTCCTCCACTCATTCTCA 1103
 Qy 302 GlyThrValArgProIleCysLeuProPheAspGluGluLeuThrProAlaThrPro 321
 Db 1104 GGCACAGTCAGGCCCATCTGTCTGCCCTTCTTGATGAGGAGCTCACTCCAGCCACCCCA 1163
 Qy 322 LeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyGlyLysMetSerAspIleLeu 341
 Db 1164 CTCCTGGATCATTTGGATGGGCTTTTACGAGCAGAAATGGAGGAAGATGTCTGACATACGT 1223
 Qy 342 LeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAlaTyrGln 361
 Db 1224 CTGCAGGGCTCAGTCCAGGTCATTGACAGCACACGGTGTGATGCAGACGATGCGTACCAG 1283
 Qy 362 GlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyValAspThrCys 381
 Db 1284 GGGGAAGTCCCGAAGAGATGATGTGTGAGGACATCCCGAAGGGGTGTGGACACCTGC 1343
 Qy 382 GlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGlyIle 401
 Db 1344 CAGGTTGACAGTGTGTGGGCCCTCATGTACCAATCTCAGGAGTGGCATGTGTGGTGGCATC 1403
 Qy 402 ValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyValTyrThrLysValSer 421
 Db 1404 GTTAGCTGGGCTATGGCTGCGGGGGCCCGAGACCCCGAGAGTATACCAAGGTCTCA 1463

Y 422 AlaTyrLeuAenTrpIleTyrAsnValTrpLysAlaGluLeu 435
 b 1464 GCCTATCTCAACTGATCTCAATGCTCTGGAAGGCTGAGCTG 1505

RESULT 14

S-10-063-735-111
 Sequence 111, Application US/10063735
 Publication No. US2003013882A1
 GENERAL INFORMATION:
 APPLICANT: Eaton, Dan L.
 APPLICANT: Filvaroff, Ellen
 APPLICANT: Gerritsen, Mary E.
 APPLICANT: Goddard, Audrey
 APPLICANT: Godowski, Paul J.
 APPLICANT: Grimaldi, Christopher J.
 APPLICANT: Gurney, Austin L.
 APPLICANT: Watanabe, Colin K.
 APPLICANT: Wood, William I.
 TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
 FILE REFERENCE: P3230R1C1
 CURRENT APPLICATION NUMBER: US/10/063,735
 CURRENT FILING DATE: 2002-05-08
 Prior Application removed - See Palm or File Wrapper
 NUMBER OF SEQ ID NOS: 170
 SEQ ID NO 111
 LENGTH: 2063
 TYPE: DNA
 ORGANISM: Homo Sapien

S-10-063-735-111

Ligment Scores:

red. No.: 7,76e-278 Length: 2063
 core: 2297,50 Matches: 429
 Percent Similarity: 98.95% Conservative: 0
 est Local Similarity: 98.85% Mismatches: 5
 Query Match: 98.10% Indels: 0
 B: 12 Gaps: 1

S-09-607-745-2 (1-435) x US-10-063-735-111 (1-2063)

Y 2 AspProAspSerAspGlnProLeuAsnSerLeuAspValLysProLeuArgLysProArg 21
 b 219 GATCTGACAGTATCAACTTGACAGCCTCGATGTCAAAACCCCTCGGCAAAACCCCGT 278
 Y 22 IleProMetGluThrPheArgLysValGlyIleProIleIleIleAlaLeuLeuSerLeu 41
 b 279 ATCCCATGGAGACCTTCAGAAAGTGGGATCCCATCATCATAGCACTACTGAGCCCTG 338
 Y 42 AlaSerIleIleIleValValValLeuIleLysValIleLeuAspLysTyrTyrPheLeu 61
 b 339 GCGAGTATCATCTGTGTGTCTCTCATCAAGGTGATTCGGATAAATACTACTTCCTC 398
 Y 62 CysGlyGlnProLeuHisPheIleProArgLysGlnLeuCysAspGlyGluLeuAspCys 81
 b 399 TSGCGGAGCCCTCTCCACTTCATCCCGAGAAGCAGCTGTGTGACGGAGAGTGGACTGT 458
 Y 82 ProLeuGlyGluAspGluGluHisCysValLysSerPheProGluGlyProAlaValAla 101
 b 459 CCCITGGGGAGGAGGAGAGACCTGTGTCAAGAGCTTCCCGGAAGGGCCCTGCAATGGCA 518
 Y 102 ValArgLeuSerLysAspArgSerThrLeuGlnValLeuAspSerAlaThrGlyAsnTrp 121
 b 519 GTCCGCTCTCCAGGACCGATCCACACTGCAAGGTGTGGACTCGGCCACAGGGNACTGG 578
 Y 122 PheSerAlaCysPheAspAsnPheThrGluAlaLeuAlaGluThrAlaCysArgGlnMet 141
 b 579 TTCTCTGCTGCTTTTCGACAACTTCAGAAAGCTCTCGCTGACAGACCTGTAGGCAGATG 638
 Y 142 GlyTyrSerSerLysProThrPheArgAlaValGluIleGlyProAspGlnAspLeuAsp 161
 b 639 GCCTACAGC-----AGAGCTGTGGAGATTGGCCACAGCAGCATCTGGAT 683

QY 162 ValValGluIleThrGluAsnSerGlnGluLeuArgMetArgAsnSerSerGlyProCys 181
 Db 684 GTTGTGAATACAGAAACAGCCAGGAGCTTCGATGCGGAACCTCAAGTGGGCCCTGT 743
 QY 182 LeuSerGlySerLeuValSerLeuHisCysLeuAlaCysGlyLysSerLeuLysThrPro 201
 Db 744 CTCTCAGGCTCCCTGGTCTCCCTGCATCTGTCTGTGGGAAAGAGCTGAAGACCCCC 803
 QY 202 ArgValValGlyGlyGluGluAlaSerValAspSerTrpProTrpGlnValSerIleGln 221
 Db 804 CGTGTGTGGTGGGAGGAGGCTCTCTGTGATTTCTTGGCTTGGCAGGTGAGCATCCAG 863
 QY 222 TyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThra 241
 Db 864 TACGACAAACAGCACAGCTCTGTGGAGGAGCATCTCGACCCCCACTGGTCTCCTCAGGCA 923
 QY 242 AlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySerAsp 261
 Db 924 GCCCAGCTGCTTCAGAAACATACCGATGTGTCAACTGGAAGGTGCGGCGCAGGCTCAGC 983
 QY 262 LysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleGluPheAsnPro 281
 Db 984 AAAGTGGGAGCTTCCCATCTCTGGCTGTGGCCAGATCATCATCATGAATTCACCCCC 1043
 QY 282 MetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPheSer 301
 Db 1044 ATGTACCCCAAGACAATGATCGCCCTCATGAAGCTGAGTTCACACTCCACTTCTCTCA 1103
 QY 302 GlyThrValArgProIleCysLeuProPheAspGluGluLeuThrProAlaThrPro 321
 Db 1104 GGCACAGTCAAGGCCATCTGTCTGCCCTTCTTGTATGAGGAGCTCACTCCAGCCACCCCA 1163
 QY 322 LeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyGlyLysMetSerAspIleLeu 341
 Db 1164 CTCTGGATCATTTGATGGGCTTTACGAGCAGATGAGGGAAGATGTCTGACATATCTG 1223
 QY 342 LeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAspAlaTyrGln 361
 Db 1224 CTGCAGGCGTCAGTCCAGGTCAATTGACAGCACACGGTGCAATGACAGCAGTGCCTAC 1283
 QY 362 GlyGlnValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThrCys 381
 Db 1284 GGGGAAGTCCCGAAGAGATGATGTGTGCGAGCATCCCGGAAGGGGTGTGACACCTGC 1343
 QY 382 GlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGlyIle 401
 Db 1344 CAGGCTGACAGTGTGTGGGCCCTTGATGTACCAATCTGACAGTGGCATGTGTGGGATC 1403
 QY 402 ValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyValTyrThrLysValSer 421
 Db 1404 GTTAGCTGGGCTATGGCTGCGGGGCCCCGAGCACCCAGAGTATACACCAAGTCTCA 1463
 QY 422 AlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu 435
 Db 1464 GCCTATCTCACTGATCTCAATGTCTGGAAGGCTGAGCTG 1505

RESULT 15

US-10-006-130A-274
 ; Sequence 274, Application US/10006130A
 ; Publication No. US20030148375A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Baker, Kevin P.
 ; APPLICANT: Botstein, David
 ; APPLICANT: Desnoyers, Luc
 ; APPLICANT: Eaton, Dan I.
 ; APPLICANT: Ferrara, Napoleone
 ; APPLICANT: Fong, Sherman
 ; APPLICANT: Gao, Wei-Qiang
 ; APPLICANT: Goddard, Audrey
 ; APPLICANT: Godowski, Paul J.
 ; APPLICANT: Grimaldi, Christopher J.
 ; APPLICANT: Gurney, Austin L.

APPLICANT: Hillan, Kenneth J.
 APPLICANT: Pan, James
 APPLICANT: Paoni, Nicholas P.
 TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
 FILE OF INVENTION: Acids Encoding the Same
 FILE REFERENCE: P2830P1C7
 CURRENT APPLICATION NUMBER: US/10/006.130A
 CURRENT FILING DATE: 2002-03-19
 Prior Application removed - See File Wrapper or Palm
 NUMBER OF SEQ ID NOS: 477
 SEQ ID NO 274
 LENGTH: 2063
 TYPE: DNA
 ORGANISM: Homo sapiens
 S-10-006-130A-274

Alignment Scores:
 red. No.: 7.76e-278 Length: 2063
 core: 2297.50 Matches: 429
 Percent Similarity: 98.85% Conservative: 0
 Best Local Similarity: 98.85% Mismatches: 0
 Query Match: 98.10% Indels: 5
 B: 12 Gaps: 1

S-09-607-745-2 (1-435) x US-10-006-130A-274 (1-2063)

y	2	AspProAspSerAspGlnProLeuAsnSerLeuAspValLysProLeuArgLysProArg	21
b	219	GATCCTGACAGTATCAACTCTTGAACAGCTCGATGTCARAACCCCTGGCGAACCCTG	278
y	22	IleProMetGluThrPheArgLysValGlyIleProLleIleLeuLeuSerLeu	41
b	279	ATCCCCATGGAGACCTTCAGAAAGTGGGATCCCATCATAGCAGTACTAGAGCTG	338
y	42	AlaSerIleIleValValValValValValValValValValValValValValVal	61
b	339	CGAGATATCATATGTTGTTCTCTCATCAAGGTGATCTGGATAAATACTACTCTCTC	398
y	62	CysGlyGlnProLeuHisPheIleProArgLysGlnLeuCysAspGlyGluLeuAspCys	81
b	399	TGCGGCGAGCTCTCCACTTCATCCCGAGGAGAGAGCTGTGTGACGGAGAGCTGGACTGT	458
y	82	ProLeuGlyGluAspGluGluHisCysValLysSerPheProGluGlyProAlaValAla	101
b	459	CCCTTGGGGAGGAGGAGAGAGCTGTGTCAAGAGCTTCCCGAAGGCGCTGCAGTGCA	518
y	102	ValArgLeuSerLysAspArgSerThrLeuGlnValLeuAspSerAlaThrGlyAsnTrp	121
b	519	GTCCGCTCTCCAGGACCGATCCACTGCAGGTGTGGACTGGCCACAGGGAAGCTGG	578
y	122	PheSerAlaCysPheAspAsnPheThrGluAlaLeuAlaGluThrAlaCysArgGlnMet	141
b	579	TTCTCTGCTCTTTCGACAACTTCACAGAAGCTCTCGCTGAGACAGAGCTGTAGCAGATG	638
y	142	GlyTyrSerSerLysProThrPheArgAlaValGluIleGlyProAspGlnAspLeuAsp	161
b	639	GGCTACAGC-----AGAGCTGTGAGATTGCCCGAGAGCTGAGATCTGGAT	683
y	162	ValValGluIleThrGluAsnSerGlnGluLeuArgMetArgAsnSerSerGlyProCys	181
b	684	GTCTGTGAATCAGAGAAACAGCAGAGAGCTTGCATGCGGAATCAAGTGGGCCCTGT	743
y	182	LeuSerGlySerLeuValSerLeuHisCysLeuAlaCysGlyLysSerLeuLysThrPro	201
b	744	CTCTCAGGCTCCCTGGTCTCCCTGCACGTCTTGCCTGTGGGAGAGAGCTGAGACCCCC	803
y	202	ArgValValGlyGlyGluGluAlaSerValAspSerTrpProTrpGlnValSerIleGln	221
b	804	CGTGTGTGGTGGGAGAGAGGCTCTGTGGATTCTTGGCCCTTGGCAGTCCAGATCCAG	863
y	222	TyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThrAla	241
b	864	TACGACAAACAGCAGCTCTGTGGAGGGAGCATCTTGGACCCCCCACTGGGTCTCTCAGCGCA	923

Search completed: November 23, 2003, 12:22:18
 Job time : 514.909 secs

Qy	242	AlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySerAsp	261
Db	924	GCCCACTGCTTCAGGAAACATACCGATGTGTCAACTGGAGGTGGCGGCTCAGAC	983
Qy	262	LysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleIleGluPheAsnPro	281
Db	984	AAACTGGGCAAGCTTCCCATCCCTGGCTGTGGCCCAAGATCATCATTTGAATCAACCCC	1043
Qy	282	MetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPheSer	301
Db	1044	ATGTACCCCAAGACAAATGATGATGCGCCCTCATGAAGCTGCAGTTCACACTTCTCA	1103
Qy	302	GlyThrValArgProIleCysLeuProPhePheAspGluGluLeuThrProAlaThrPro	321
Db	1104	GGCACAGTCAGGCCCATCTGTCTGCCCTTCTTTGATGAGGAGCTCACTCCAGCCACCCCA	1163
Qy	322	LeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyGlyLysMetSerAspIleLeu	341
Db	1164	CTCTGGATCATTTGGATGGGCTTTTACGACAGATGGAGGAAGATGTCTGACATATCTG	1223
Qy	342	LeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAlaTyrGln	361
Db	1224	CTGAGCGTCAGTCCAGGTCATTGACAGCACAGCGTGCAATGCAGACGATGCTACCA	1283
Qy	362	GlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThrCys	381
Db	1284	GGGGAAGTCACCGAGAAGATGATGTGTCAGGATATCCCGAAGGGGTGTGGACACCTGC	1343
Qy	382	GlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGlyIle	401
Db	1344	CAGGTCACAGTGTGGGCGCCCTGATGTACCAATCTGACCAAGTGGCATGTGTGGGATC	1403
Qy	402	ValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyValTyrThrLysValSer	421
Db	1404	GTTAGCTGGGGCTATGGCTCGGGGGCCCGAGACCCCGAGAGTATACACCAAGGTCTCA	1463
Qy	422	AlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu	435
Db	1464	GCCTATCTCAACTGGATCTTACAATGTCTGGAAGGCTGAGCTG	1505

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

1 protein - nucleic search, using frame_plus_p2n model

in on: November 23, 2003, 07:53:00 ; Search time 3092.27 Seconds
(without alignments)
3418.993 Million cell updates/sec

File: US-09-607-745-2
Effect score: 2342
Sequence: 1 MDPDSQPLNSLDVKEPLRKP.....VYTKVSAYLNWYNNWKAEL 435

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Archived: 22781392 seqs, 12152238056 residues

Real number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Maximum Match 0%

Listing first 45 summaries

Command line parameters:

MODEL=frame+ p2n.model -DEV=xlh
p=/cgn2.1/USPTO.spool/US09607745/runat 21112003 144346 22229/app query.fasta_1.1038
DB=EST -QFMT=fastap -SUFFIX=est -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
LOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
NITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=200000000
USER=US09607745 @CGN 1.1 3596 @runat 21112003 144346 22229 -NCFU=6 -ICFU=3
NO MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSFLOCK=100 -LONGLOG
REV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
XGAPEXT=7 -XGAPOP=10 -XGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST.*
1: em_estba.*
2: em_esthum.*
3: em_estin.*
4: em_estnu.*
5: em_estov.*
6: em_estpl.*
7: em_estro.*
8: em_htc.*
9: gb_est1.*
10: gb_est2.*
11: gb_htc.*
12: gb_est3.*
13: gb_est4.*
14: gb_est5.*
15: em_estfun.*
16: em_eston.*
17: em_gss_hum.*
18: em_gss_inv.*
19: em_gss_pln.*
20: em_gss_vrt.*
21: em_gss_fun.*
22: em_gss_mam.*
23: em_gss_mus.*
24: em_gss_pro.*
25: em_gss_rod.*
26: em_gss_phg.*
27: em_gss_vrl.*
28: gb_gss1.*

29: gb_gss2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Length	DB	ID	Description
1	2044	87.3	1973	11	BC004855	BC004855 Homo sapi
2	1668.5	71.2	2177	11	AK078890	AK078890 Mus muscu
3	1233	52.6	844	10	BE531100	BE531100 601278486
4	1203	51.4	895	13	BU522841	BU522841 AGENCOURT
5	1158.5	49.5	975	10	BG288427	BG288427 602388091
6	1108.5	47.3	920	10	BG616663	BG616663 601278696
7	1054.5	45.0	617	12	BM795149	BM795149 K-EST0076
8	1042	44.5	956	13	BU523218	BU523218 AGENCOURT
9	971	41.5	569	12	BM686292	BM686292 UI-E-CRO-
10	951.5	40.6	804	12	BG822914	BG822914 602727917
11	949.5	40.5	1007	10	BG386903	BG386903 602454702
12	893	38.1	920	13	BQ322656	BQ322656 AGENCOURT
13	891	38.0	689	12	BG966811	BG966811 602834306
14	890.5	38.0	879	10	BG616186	BG616186 601278758
15	885	37.8	515	14	CB142902	CB142902 K-EST0196
16	881.5	37.6	645	10	BF118991	BF118991 601755351
17	877	37.4	503	10	BE514663	BE514663 601317126
18	875.5	37.4	787	10	BE615750	BE615750 601279885
19	858	36.6	498	12	BM795128	BM795128 K-EST0076
20	856	36.5	548	14	CB159713	CB159713 K-EST0219
21	848	36.2	744	12	BM008902	BM008902 603618532
22	844	36.0	742	10	BG468475	BG468475 602510448
23	832.5	35.5	693	10	BG481239	BG481239 602528566
24	825.5	35.2	935	10	BE615106	BE615106 601280475
25	816	34.8	659	9	AI924537	AI924537 wn61c07.x
26	808	34.5	648	9	AI924132	AI924132 wn53f11.x
27	808	34.5	931	13	BUL57248	BUL57248 AGENCOURT
28	797	34.0	563	12	BM773306	BM773306 K-EST0057
29	797	34.0	572	12	BM746729	BM746729 K-EST0021
30	797	34.0	578	12	BM747250	BM747250 K-EST0021
31	797	34.0	593	12	BM788452	BM788452 K-EST0067
32	797	34.0	596	12	BM772743	BM772743 K-EST0056
33	797	34.0	614	12	BM788163	BM788163 K-EST0067
34	797	34.0	633	12	BM764659	BM764659 K-EST0046
35	797	34.0	691	12	BM763697	BM763697 K-EST0045
36	797	34.0	692	12	BM772754	BM772754 K-EST0056
37	797	34.0	700	12	BM772620	BM772620 K-EST0056
38	783	33.4	548	12	BM746721	BM746721 K-EST0021
39	775.5	33.1	977	10	B614660	B614660 601281614
40	770	32.9	462	12	BG984172	BG984172 IL5-CN006
41	765.5	32.7	1088	10	BE616293	BE616293 601281476
42	746.5	31.9	1835	11	BC048135	BC048135 Danilo rer
43	740.5	31.6	1105	10	BE736425	BE736425 601307316
44	737	31.5	850	12	BT251465	BT251465 602994002
45	728	31.1	803	14	CA311909	CA311909 UI-CF-FNO

ALIGNMENTS

RESULT 1
BC004855
LOCUS
DEFINITION Homo sapiens, Similar to transmembrane protease, serine 4, clone
IMAGE:3835263, mRNA.
ACCESSION BC004855
VERSION BC004855.1 GI:14709533
KEYWORDS HTC.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1973)

AUTHORS TITLE JOURNAL

Strausberg, R.
Direct Submission
Submitted (21-MAR-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgapbs-remail.nih.gov
Tissue Procurement: DCTD/DTP
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed By: The I.M.A.G.E. Consortium (ILNL)
DNA Sequencing By: Institute for Systems Biology
<http://www.systemsbio.org>
contact: amadams@systemsbio.org
Anup Madan, Rachel Dickhoff, Jessica Fahey, Stephanie Ford, Julia
Greene, Mark Kettelman and Anuradha Madan

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/ILNL at: <http://image.llnl.gov>
Series: IRAL Plate: 14 Row: p Column: 19
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 8347148
This clone has the following problem: frame shifted.

FEATURES

Location/Qualifiers

1..1973
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:3835263"
/tissue_type="Ovary, adenocarcinoma"
/clone_lib="NIH MGC.9"
/lab_host="DH10B-R"
/note="Vector: pOTB7"

ASE COUNT 471 a 573 c 531 g 398 t

ALIGNMENT

Alignment Scores:

red. No.: 2,51e-210 Length: 1973
core: 2044.00 Matches: 387
percent Similarity: 88.76% Conservative: 0
est Local Similarity: 88.76% Mismatches: 0
very Match: 87.28% Indels: 49
B: 11 Gaps: 1

S-09-607-745-2 (1-435) x BC004855 (1-1973)

1 MetAspProAspSerAspGlnProLeuAsnSerLeuAspValLysProLeuArgLysPro 20
243 ATGGATCCTGACGTGATCAACCTCTGACACGCTCGATGTCACACCCCTCGCAACCC 302
21 ArgTleProMetGluThrPheArgLysValGlylleProIlelleAlaLeuLeuSer 40
303 CGTATCCCATGGAGACCTTCAGAAAGTGGGATCCCATCATCATAGCACTACTGAGC 362
41 LeuAlaSerllellelleValValValleulleLysVallelleLeuAspLysTyrPhe 60
363 CTGGCGAGTATCATATTGTGTGCTCTCATCAAGTGATTCGGATAAATACTACTTC 422
61 LeuCysGlyGlnProLeuHisPheIleProArgLysGlnLeuCysAspGlyGluLeuAsp 80
423 CTCGCGGGCGACCTCTCCACTTCATCCGAGGAGACGCTGTGTGACGGAGCTGGAC 482
81 CysProLeuGlyGluAspGluHisCysValLysSerPheProGluGlyProAlaVal 100
483 TGTCCCTTGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 542
101 AlaValArgLeuSerLysAspArgSerThrLeuGlnValLeuAspSerAlaThrGlyAsn 120
543 GCAGTCCGCGCTCTCCAGAGGACCGATCCACATCGCAGGTGTGAGCTCGGCCACAGGAA 602
121 TrpPheSerAlaCysPheAspAsnPheThrGluAlaLeuAlaGluThrAlaCysArgGln 140

603 TGGTCTCTGCTGCTGTTTCGACCAACTTTCACAGAAAGCTCTCGTGTGAGACAGCCTGTAGCAG 662
141 MetGlyTyrSerSerLysProThrPheArgAlaValGluIleleGlyProAspGlnAspLeu 160
663 ATGGGCTACAGCAGCAAAACCCACTTTCAGAGCTGTGTGAGATTGGCCCGACAGGATCTG 722
161 AspValValGluIleThrGluAsnSerGlnGluLeuArgMetArgAsnSerSerGlyPro 180
723 GATGTTGTTTGAATACAGAAACAGCCAGAGGCTTCGATGCGGAACCTCAAGTGGGCC 782
181 CysLeuSerGlySerLeuValSerLeuHisCysLeuAlaCysGlyLysSerLeuLysThr 200
783 TGTCTCTCAGGCTCCCTGTGCTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 842
201 ProArgValValGlyGlyGluGluAlaSerValAspSerTrpProTrpGlnValSerIle 220
843 CCCGCTGTGTGGTGGGAGGAGGCTCTGTGTGATTCTTGGCTTGGCAGTGCAGATC 902
221 GlnTyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThr 240
903 CAGTACGACAAAACAGCAGCAGCTCTGTGAGGAGGAGCATCTCGACCCCTGGTCTCTCAG 962
241 AlaAlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySer 260
963 GCAGCCACTGCTTCAGGAAACATACCGATGTTTCACTGGAGGTGGCGGAGGCTCA 1022
261 AspLysLeuGlySerPheProSerLeuAlaValAlaLysIlellelleGluPheAsn 280
1023 GACAAACTGGGAGCTTCCCATCCCTGGCTGTGGCAAGATCATCATTAATGAATCAAC 1082
281 ProMetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPhe 300
1083 CCCATGTACCCCAAGACAAATGACATCGCCCTCATGAAGTGCAGTTCACACTCCTTC 1142
301 SerGlyThrValArgProIleCysLeuProPheAspGluGluLeuThrProAlaThr 320
1143 TCAGGCACAGTCAGGCCCATCTGTCTGCCCTCTTTGATGAGGAGCTCATCTCAGCCACC 1202
321 ProLeuTrpIlelleGlyTrpGlyPheThrLysGlnAsnGlyLysMetSerAspIle 340
1203 CCATCTCGATCATTTGATGGGCTTTACGAAGCAGAAATGGA----- 1244
341 LeuLeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAspAlaTyr 360
1244 ----- 1244
361 GlnGlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThr 380
1244 ----- 1244
381 CysGlnGly-AspSerGlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValValG 400
1245 -----GGGTGACAGTGGTGGGCCCTGTGATGTACCAATCTCACCAGTGGCATGTGTGG 1298
400 YlleValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyValThrLysVa 420
1299 CATCGTTAGTGGGCTATGGCTGGGGGGCCCGAGCACCCAGGAGATATACACCAAGGT 1358
420 lSerAlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu 435
1359 CTCAGCCTATCTCACTGGATCTACATATGTCGGAAGGCTGAGCTG 1404

RESULT 2

AK078890

LOCUS

DEFINITION

AK078890

ACCSSION

VERSION

KEYWORDS

SOURCE

ORGANISM

AK078890 2177 bp mRNA linear HTC 05-DEC-2002
Mus musculus adult male colon cDNA, RIKEN full-length enriched
library, clone:9030622G02 product:SIMILAR TO TRANSMEMBRANE
PROTEASE, SERINE 4 homolog [Mus musculus], full insert sequence.
AK078890.1 GI:26098158
HTC; CAP trapper.
Mus musculus (house mouse)
Mus musculus


```

Y 120 nTrpPheSerAlaCysPheAspAsnPheThrGluAlaLeuAlaGluThrAlaCysArgGI 140
b 559 CTGGCCCTCACTCTGTTTCGACAACTTACAGAGAGCAGTGGCCAGAGCAGCGCTCAGACA 618
Y 140 nMetGlyTyrSerSerLysProThrPheArgAlaValGluLeuGlyProAspGlnAspLe 160
b 619 GATGGGTATACAGACCCGCGCTTTCTCGCTGCTTTCAGAGCAGTGGAGATCCGTCAGATCAGACCT 678
Y 160 uAspValValGluLeuThrGluAsnSerGlnGluLeuArgMetArgAsnSerSerGlyPr 180
b 679 CCCTGTGCTCAAGTCACAGAAACAGCAGCAACTTCAGTGCAGATGGAAGCAGATC 738
Y 180 cCysLeuSerGlySerLeuValSerLeuHisCysLeuAlaCysGlyLysSerLeuLysTh 200
b 739 CTGCCTCTCAGGCTCCCTGGTTCTCTCGCTGCTTCTGACTGTGGAAGAGCCTGAAGAC 798
Y 200 rProArgValValGlyGlyGluGluAlaSerValAspSerTrpProThrGlnValSerIl 220
b 799 TCCTCGTGTGGTGGGTGGAGGCCCTGTGGATTCTTTGGCGGTGGCAGGTCCAGCAT 858
Y 220 eGlnTyrAspLysGlnHisValCysGlySerLeuLeuAspProHisTrpValLeuTh 240
b 859 CCAGTACACAGCAGCAGTGTCTGTGGTGGAGCATCTCTGGATCCCACTGATCCTCAC 918
Y 240 rAlaAlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySe 260
b 919 AGCAGCCCACTGCTTCAGAAAGTATCTGTGTGTCAGCTGGAAGGTTCAGGCGAGGCTC 978
Y 260 rAspLysLeuGlySerPheProSerLeuAlaValAlaLysLeuLeuLeuGluPheAs 280
b 979 AAACATCTGGGTAACTCTCCATCTCTCCCTGGCCCAAGATCTTCATCGCTGAACCCAA 1038
Y 280 nProMetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPh 300
b 1039 TCCTCTGTACCCCAAGAGAGGACATTGCCCTTGTAAAGTGGCAGATGCCACTCACATT 1098
Y 300 eSerGlyThrValArgProLysCysLeuProPhePheAspGluLeuThrProAlaTh 320
b 1099 CTCAGGCTCAGTCAGGCCCATCTGCTGCCCTTCTCTGATGAGGTGCTTGTCCAGGCCAC 1158
Y 320 rProLeuTrpIleLeuGlyTrpGlyPheThrLysGlnAsnGlyLysMetSerAspIl 340
b 1159 ACCAGTCTGGGTCACTGATGGGCTTTTACAGAGAAACCGAGAAAGATGCTTGACAT 1218
Y 340 eLeuLeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAspAlaTy 360
b 1219 GCTACTGCAGGCATCAGTCCAGGTCAATTGACAGCACACGGTGCATGCAGAGGATGCCTA 1278
Y 360 rGlnGlyGluValThrGluLysMetMetCysAlaGlyLeuProGluGlyGlyValAspTh 380
b 1279 CGAAGGGGAAGTGACCGCTGAGATGCTGTGTGCAGGTACCCACAGGGGTGGCAAGGACAC 1338
Y 380 rCysGlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGl 400
b 1339 CTGCAGGGTGACAGTGGTGGCCCTTTGATGATCCATTCTGACAGTGGCAGGTAGTAGG 1398
Y 400 yIleValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyValTyrThrLysVa 420
b 1399 CATCGTAGCTGGGGCCATGATGCGGGGCCCAAGTACTCTCGAGGTGATATACCAAGGT 1458
Y 420 lSerAlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu 435
b 1459 CACTGCCTATCTCACTGATGATCTACATGTTTCGGAAGTCTGAGATG 1504

```

```

RESULT 3
E531100
OCUS BE531100 844 bp mRNA linear EST 09-AUG-2000
EFINITION 601278466F1 NIH_MGC_39 Homo sapiens cDNA clone IMAGE:3610337 5',
mRNA sequence.
CCESION BE531100
ERSTION BE531100.1 GI:9759745
EYWORDS EST.
SOURCE Homo sapiens (human)

```

```

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 844)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: gcapbs-i@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LCM266 row: d column: 18
High quality sequence stop: 712.
Location/Qualifiers
source 1..844
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:3610337"
/tissue_type="adenocarcinoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH MGC 39"
/note="Organ: pancreas; Vector: pOTB7; Site: 1: XhoI;
Site 2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCAGAG(G). Library constructed
by Ling Hong in the laboratory of Gerald M. Rubin
(University of California, Berkeley) using ZAP-cDNA
synthesis kit (Stratagene) and Superscript II RT (Life
Technologies)."
BASE COUNT 165 a 251 c 244 g 184 t
ORIGIN

```

Alignment Scores:

```

Pred. No.: 6,12e-123 Length: 844
Score: 1233.00 Matches: 254
Percent Similarity: 91.49% Conservative: 4
Best Local Similarity: 90.07% Mismatches: 14
Query Match: 52.65% Indels: 15
DB: 10 Gaps: 2

```

US-09-607-745-2 (1-435) x BE531100 (1-844)

```

Qy 40 SerLeuAlaSerIleIleValValValLeuLeuLysValIleLeuAspLysTyrTyr 59
Db 1 AGCCTGGCGAGTATCATCATTTGTGTCTCATCAAGGTGATTCGGATAAATACTAC 60
Qy 60 PheLeuCysGlyGlnProLeuHisPheIleProArgLysGlnLeuCysAspGlyGluLeu 79
Db 61 TTCCTCTGCGGGCAGCTCTCCACTTCATCCCGAGGAGCAGCTGTGTGACGAGAGCTG 120
Qy 80 AspCysProLeuGlyGluAspGluGluHisCysValLysSerPheProGluGlyProAla 99
Db 121 GACTGTCCCTTGGGGAGGAGCAGGAGCACTGTGTCAAGAGCTTCCCGAAGGCGCTGCA 180
Qy 100 ValAlaValArgLeuSerLysAspArgSerThrLeuGlnValLeuAspSerAlaThrGly 119
Db 181 GTGGCAGTCCGCTCTCCAGAGCCGATCCACACTCGCAGGTGTGTGACTCGGCCACAGGG 240
Qy 120 AsnTrpPheSerAlaCysPheAspAsnPheThrGluAlaLeuAlaGluThrAlaCysArg 139
Db 241 AACTGGTCTCTGCTGTTTCGACAACTTCACAGAGCTCTCGCTGAGCAGGCTGTAGG 300
Qy 140 GlnMetGlyTyrSerSerLysProThrPheArgAlaValGluLeuGlyProAspGlnAsp 159
Db 301 CAGATGGGCTACAGCAGCAAAACCACCTTTTCAGAGCTGTGGAGATTGGCCAGACAGGAT 360
Qy 160 LeuAspValValGluLeuThrGluAsnSerGlnGluLeuArgMetArgAsnSerSerGly 179

```

```

> 361 CTGGATGTTTGAATACAGAAAACAGCAGAGCTTCGATCGGAACCTCAAGTGGG 420
/ 180 ProCysLeuSerGlySerLeuValSerLeuHisCysLeuAlaCysGlyLysSerLeuLys 199
> 421 CCTGTCTCTCAGGCTCCCTGGTCTCCCTGCACCTGTCTTGCCTGTGGAGAGCCTGAAG 480
/ 200 ThrProArgValValGlyGlyGluGluAlaSerValAspSerTrpProTrpGlnValSer 219
> 481 ACCCCCGCTGTGGTGGGAGAGGCGCTCTGTGGATCTTGGCTTGGCAGTCAAG 540
/ 220 IleLeuTrpAspLysGlnHisValCysGlyGlySerLeuAlaSerProHisTrpValLeu 239
> 541 ATCCAGTACACAAACAGCAGCTCTGTGGAGGAGCATCTGGACCCCTGGTCTC 600
/ 240 ThrAlaAlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGly 259
> 601 ACGGCAGC-CACTGTCTCAGAAACATACCGATGTGTCACTGGAGGTGGGGCAGGC 659
/ 260 SerAspLysLeuGlySerPheProSerLeuAlaValAlaLysLeuIleIleGluPhe 279
> 660 TCAGACAA-CTGGGCAGCTTCCCATCCCTGGGCTGTGGCCAGCATCATCATTTGAATCC 718
/ 280 AspProMetTrpProLysAspAsnAspIleAlaLeuMetLysLeuGlnPhePro----- 297
> 719 AACCTGT---CCCCAAGACAT-GACATCGCCCTCATGAAGCTGCG-TTCCCTCTCACTT 773
/ 298 -----LeuThrPheSerGlyThrValArgProLysLeuPro 310
> 774 CTCGGCCCGCGGCTCTGTGTCTTGTGGGGGCCCT-CGCCCCCTCTGTCTGGGG 832
/ 311 PhePhe 312
> 833 GGGTTC 838

RESULT 4
> J522841
> DEFINITION
> B522841
> ACESION
> B522841
> XNAMES
> B522841.1 GI:22833279
> EST.
> Mus musculus (house mouse)
> Mus musculus
> Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
> Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
> NIH-MGC http://mgs.nci.nih.gov/.
> National Institutes of Health, Mammalian Gene Collection (MGC)
> Unpublished
> Contact: Robert Strausberg, Ph.D.
> Email: gcapbs-remail.nih.gov
> Tissue Procurement: The Cepko Laboratory
> CDNA Library Preparation: Life Technologies, Inc.
> CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LIML)
> DNA Sequencing by: Agencourt Bioscience Corporation
> Clone distribution: MGC clone distribution information can be
> found through the I.M.A.G.E. Consortium/LIML at:
> http://image.llnl.gov
> Plate: L14M4129 row: i column: 06
> High quality sequence stop: 665.
> Location/Qualifiers
> 1. .895
> /organism="Mus musculus"
> /mol_type="mRNA"
> /strain="FVB/N"
> /db_xref="taxon:10090"
> /clone="IMAGE:6529422"
> /lab_host="DH10B (T1 phage-resistant)"
> /clone_lib="NCI CGAP Co24"
> /note="Organ: colon; Vector: pCMV-SPORT6; Site_1: NotI;
> Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt.

FEATURES
source

```

```

Average insert size 1.6 kb. Constructed by Life
Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT      202 a      242 c      257 g      194 t
ORIGIN

Alignment Scores:
Pred. No.:      1.2e-119      Length:      895
Score:          1203.00      Matches:      217
Percent Similarity: 89.17%      Conservative: 30
Best Local Similarity: 78.34%      Mismatches: 29
Query Match:      51.37%      Indels:      1
DB:              13          Gaps:      0

US-09-607-745-2 (1-435) x B522841 (1-895)

Qy 154 IleGlyProAspGlnAspLeuAspValValGluIleThrGluAsnSerGlnGluLeuArg 173
Db 23 GTCCGTCCAGATCAGAACCTCCCTGTGTCTCAAGTCACAGAAACAGCCAGGAACTTCAG 82
Qy 174 MetArgAsnSerSerGlyProCysLeuSerGlySerLeuValSerLeuHisCysLeuAla 193
Db 83 GTCAGATATGGAAGCAGATCCCTGCTCAGGCTCCCTGCTTCCCTTGCCTGCCTTGAC 142
Qy 194 CysGlyLysSerLeuLysThrProArgValValGlyGlyGluGluAlaSerValAsnSer 213
Db 143 TGTGGAAAGAGAGCTGAAGACTCCTGCTGTGGTGGGTGGAGGCCCTGTGGATCTT 202
Qy 214 TrpProTrpGlnValSerIleGlnTrpAspLysGlnHisValCysGlyGlySerIleLeu 233
Db 203 TGGCCCTGGCAGGTACGATCCAGTACAAACAGCAGCATGTCTGTGTGGAGCATCTTG 262
Qy 234 AspProHisTrpValLeuThrAlaAlaHisCysPheArgLysHisThrAspValPheAsn 253
Db 263 GATCCCACTGGATCCTCAGCAGGCCACTGCTTCAGGAAGTATCTTGATGTGTCAAGC 322
Qy 254 TrpLysValArgAlaGlySerAspLysLeuGlySerPheProSerLeuAlaValAlaLys 273
Db 323 TGAAGGTGAGGAGCAGGCTCAAACTACTGGGTAACCTCCATCCTTGCCTGTGGCCAG 382
Qy 274 IleIleIleIleGluPheAsnProMetTrpProLysAspAsnAspIleAlaLeuMetLys 293
Db 383 ATCTTCATCGCTGAACCCCAATCCTCTGTATCCCCCAAGAGAGGACATTCCTTGTAA 442
Qy 294 LeuGlnPheProLeuThrPheSerGlyThrValArgProLysCysLeuProphePheAsp 313
Db 443 CTGAGATGCCACTCATTCTCAGGCTCAGTCAGGCCCATCTGCTGCTGCTTCTCTGAT 502
Qy 314 GluGluLeuThrProAlaThrProLeuTrpIleIleGlyTrpGlyPheThrLysGlnAsn 333
Db 503 GAGGTGCTTGTCCCGCCACACCCAGTCTGGGTCAATTGGATGGGGCTTTACAGAGAAAC 562
Qy 334 GlyGlyLysMetSerAspIleLeuLeuGlnAlaSerValGlnValIleAspSerThrArg 353
Db 563 GGAGGAAAGATGCTGCACATGTCTACTGCGGCATCAGTCCAGGTCACTGCAGCAGCAG 622
Qy 354 CysAsnAlaAspAlaTrpGlnGlyGluValThrGluLysMetMetCysAlaGlyIle 373
Db 623 TGCATATGCAGAGGATGCTTACGAAGGGAAGTACCGCTCAGATGCTGTGTGAGGTACC 682
Qy 374 ProGluGlyGlyValAspThrCysGlnGlyAspSerGlyGlyProLeuMetTrpGlnSer 393
Db 683 CCACAGGTGGGCAAGGACACCTGCCAGGTGACAGTGGTGGGCTTTGATGTACCATCT 742
Qy 394 AspGlnTrpHisValValGlyIleValSerTrpGlyTrpGlyCysGlyGlyProSerThr 413
Db 743 GACAAGTGGCAGGTAGTAGCATCTGAGCTGGGGCCATGGATCGCGGCGGCCCAAGTACT 802
Qy 414 -ProGlyValTrpThrLysValSerAlaTrpLeuAsnTrpIleTrpAsn 429
Db 803 CCCTGGAGTGTATACCAAGGTCACTGCTATCTCACTCACTGGATCTCAAT 851

RESULT 5
BG288427

```

```

JCUS      BG288427      975 bp      mRNA      linear      EST 21-FEB-2001
DEFINITION 602388091F1 NIH_MGC_93 Homo sapiens cDNA clone IMAGE:4516990 5',
mRNA sequence.
ACCESSION  BG288427
VERSION     BG288427.1 GI:13043459
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 975)
AUTHORS     NIH-MGC http://mgc.nci.nih.gov/.
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Tissue Procurement: ATCC
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LHAMi0409 row: e column: 23
            High quality sequence stop: 659.
            Location/Qualifiers
                1..975
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /clone="IMAGE:4516990"
                /tissue_type="transitional cell papilloma, cell line"
                /lab_host="DH10B (phage-resistant)"
                /clone_lib="NIH_MGC_93"
                /note="Organ: bladder; Vector: pCMV-SPORT6; Site:1: NotI;
                Site:2: SalI; Cloned unidirectionally; oligo-dr primed.
                Average insert size 1.7 kb. Library enriched for
                full-length clones and constructed by Life Technologies.
                Note: this is a NIH MGC Library."
FEATURES             208 a 270 c 300 g 197 t
BASE COUNT          208 a 270 c 300 g 197 t
ORIGIN
lignment Scores:
red. No.:          9.23e-115      Length:          975
core:              1158.50        Matches:         221
percent Similarity: 96.94%        Conservative:    1
best Local Similarity: 96.51%     Mismatches:     4
very Match:        49.47%        Indels:         4
3:                 10             Gaps:           1
3-09-607-745-2 (1-435) x BG288427 (1-975)
Y 196 LysSerLeuLysThrProArgValValGlyGlyGluAlaSerValAspSerTrpPro 215
C 3 AAGAGCCTGAGACCCCGCTGGTGGGGAGGAGGCTCTGGGATCTTGGCCT 62
Y 216 TrpGlnValSerIleGlnTyrAspLysGlnHisValCysGlyGlySerIleLeuAspPro 235
C 63 TGGCAGGTGAGATCCAGTACGACAAACAGACAGCTCTGTGGAGGAGCATCTGGACCCC 122
Y 236 HisTrpValLeuThrAlaAlaHisCysPheArgLysHisThrAspValPheAsnTrpLys 255
C 123 CACTGGGTCTCTACGGGAGGCCACTGCTGTAGGAACATACCGATGTGTTCACTGGAAG 182
Y 256 ValArgAlaGlySerAspLysLeuGlySerPheProSerLeuAlaValAlaLysIle 275
C 183 GTGGGGGAGGCTCAGACAACTGGGAGGTTCCTATCCCTGGCTGGCCCAAGATCATC 242
Y 276 IleIleGluPheAsnProMetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGln 295
C 243 ATCATTTGATTCACCCCATGATCCCAAGACAAATGACATCGCCCTCATGAAGCTCGAG 302
Y 296 PheProLeuThrPheSerGlyThrValArgProIleCysLeuProPhePheAspGluGlu 315

```

```

Db 303 TTCCCACTCACCTTCTCAGGCACAGTCAGGCCCATCTGTCTGCCCTCTTTGATGAGGAG 362
QY 316 LeuThrProAlaThrProLeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyGly 335
Db 363 CTCACCTCCAGCCACCCCACTCTGGATCATTTGGATGGGGCTTTTACGAAGCAGATGAGGG 422
QY 336 LysMetSerAspIleLeuLeuGlnAlaSerValGlnValIleAspSerThrArgCysAsn 355
Db 423 AGATGTCCTGACATACCTGCTGAGGGCTGCTCCAGCTCATTCACACACACGGTGCAT 482
QY 356 AlaAspAlaLysTrpGlnGlyGluValThrGluLysMetMetCysAlaGlyIleProGlu 375
Db 483 GCAGACGATCGGTACCGAGGGGAGTCCCGAGAGATGATGTGTGACGATCCCGGAA 542
QY 376 -GlyGlyValAspThrCysGlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspG1 395
Db 543 GGGGGGTGTGGACACCTGCCA-GGTGACAGTGGTGGGCCCTGATGTACCAATCTGACCA 601
QY 395 nT-rpHisValValGlyIleValSerTrpGlyTrpGlyCysGlyGlyProSerThrProG1 415
Db 602 GTGGCATGTGGTGGGCATGCTTAGCTGGGGCTATGCTGGGGGCC---GAGCACCAGG 658
QY 415 yValTyr-ThrLysValSerAla 422
Db 659 AGTTTACCACCAAGGTCTCAAGC 681
RESULT 6
BE616663      920 bp      mRNA      linear      EST 24-AUG-2000
LOCUS        601278696F1 NIH_MGC_39 Homo sapiens cDNA clone IMAGE:3610990 5',
mRNA sequence.
ACCESSION    BE616663
VERSION      BE616663.1 GI:9898262
KEYWORDS     EST.
SOURCE       Homo sapiens (human)
ORGANISM     Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE     1 (bases 1 to 920)
AUTHORS       NIH-MGC http://mgc.nci.nih.gov/.
TITLE         National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL       Unpublished
COMMENT       Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Ling Hong/Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLC267 row: o column: 23
            High quality sequence start: 5
            High quality sequence stop: 819.
            Location/Qualifiers
                1..920
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /clone="IMAGE:3610990"
                /tissue_type="adenocarcinoma"
                /lab_host="DH10B (phage-resistant)"
                /clone_lib="NIH_MGC_39"
                /note="Organ: pancreas; Vector: pOTB7; Site:1: XhoI;
                Site:2: EcoRI; cDNA made by oligo-dr priming.
                Directionally cloned into EcoRI/XhoI sites using the
                following 5' adaptor: GGCACGAG(G). Library constructed
                by Ling Hong in the laboratory of Gerald M. Rubin
                (University of California, Berkeley) using ZAP-cDNA
                synthesis kit (Stratagene) and Superscript II RT (Life
                Technologies)."
FEATURES             205 a 256 c 203 t
BASE COUNT          205 a 256 c 203 t

```

21GIN

Alignment Scores:

red. No.:	2.23e-109	Length:	920
core:	1108.50	Matches:	255
percent Similarity:	86.27%	Conservative:	9
best Local Similarity:	83.33%	Mismatches:	27
very Match:	47.33%	Indels:	20
3:	10	Gaps:	5

5-09-607-745-2 (1-435) x BE616663 (1-920)

39 LeuSerLeuAla-SerIleIleValValValLeuIleLysValIleLeuAspLysTy 58
:::
15 ATAGCGCTGGCGAGGTATCATATTGTTGGTTCCTCATCAAGTGATTCGTGATAATA 74

58 rTyrPheLeuCysGlyGlnProLeuHisPheIleProArgLysGlnLeuCysAspGlyGl 78
CTACTTCCTCTGGGGCAGGCTCTCACCTTCATCCCGAGGAACAGACTGTGTACGGAGA 134
75 uLeuAspCysProLeuGluAspGluGluHiSCysValLysSerPheProGluGlyPr 98
GCTGGACTGTCCCTGGGGGAGGACGAGGAGCACTGTGTCAAGAGCTTCGCCAAGGGCC 194
98 oAlaValAlaValArgLeuSerLys-AspargSerThrLeuGlnValLeuAsnSerAlat 118
195 TGCAGTGGCAGTCGGCTCTCCAAGGAGCGATCCACACTGCAGGTGCTGAGACTGGGCCA 254
118 hrGlyAsnTrpPheSerAlaCysPheAspAsnPheThrGluAlaLeuAlaGluThrAlaC 138
255 CAGGAACTGGTCTCTGCTGTTTCGACAACCTTCAGAAAGCTCTGCCTGAGACAGCCT 314
138 ysArgGlnMetGlyTyrrSerSerLysProThrPheArgAlaValGluIleGlyProAspG 158
315 GTAGGCAGATGGGCTACAGCAGCAAACCCACTTTCAGAGCTGTGGAGATTGGCCAGACC 374
158 lnAspLeuAspValValGluIleThrGluAsnSerGlnGluLeuArgMetArgAsnSerS 178
375 AGGATCTGGAGTTGTTGAAATCACAGAAAACAGCCAGGAGCTTCGATCGGAATCTCAA 434
178 erGlyProCysLeuSerGlySerLeuValSerLeuHisCysLeuAlaCys-GlyLysSer 197
435 GTGGGCCCTCTCTCAGGCTCCCTGTCTCCCCTGCACCTGTCTGCCTGTGGGAAGAGC 494
198 LeuLysThrProArgValValGlyGlyGluAlaSerValAspSerTrpPro-TrpGl 217
495 CTGAAGAACCCCGTGGTGGTGGGGAGGAGGCCCTCTGTGGAATCTTGGCCTTTGGCA 554
217 nValSerIleGlnTyrr-AspLysGlnHisValCysGlyGlySerIleLeuAspProHist 237
555 GGTCAGCATCCAGTACGACAAACAGCACCTCTGTGGAGGGAGCATCTCGA-CGCCACT 613
237 rpValLeuThrAlaAlaHisCysPheArgLysHisThrAspValPheAsn---TrpLysV 256
614 GGGTCTCACGGH-AGGCACCTGTTCCGGAACATACCGAATGTTGTTCACACTTGAAGG 672
256 al-ArgAlaGlySer-Asp---LysLeuGlySer-----PheProSerLeuAlaValAl 272
673 TGCCGGCAGCTCAGACAAAAACCTGGGCAAGGTTCCAATACCTGGTCTGTGGGCCAA 732
272 aLysIleIleIleGluPheAsnProMetTyrrProLysAspAsn-AspIleAlaLeuM 292
733 AGATCCTTCATTAITGAATTCACCCCATGTATCCCCAAAGACAATTGACATCGCCCTCA 792
292 etLysLeuGlnPheProLeuThrPheSerGlyThrValArgProIleCysLeuProPheP 312
793 TGAGCTGAAGTT-CCACTCACTTCTCAGCACACAGCAGGCCATC-TGTTTGCT---C 847
312 heAspGluGluLeuThrProAlaThrProLeuTrpIleIleGlyTrpGlyPheThrLysG 332
848 TTGAATGAGGATC-CATCCAGCACCCCATCTGGTTCATGGTT-----GGGTTTACAGCC 900
332 In 332

Db 901 AG 902

RESULT 7

BM795149	617 bp	mRNA	linear	EST 05-MAR-2002		
LOCUS	X-EST0076707	S21SNU520	Homo sapiens	cDNA clone S21SNU520-27-H07 5'		
DEFINITION	mRNA sequence.					
ACCESSION	BM795149					
VERSION	BM795149.1	GI:19143381				
KEYWORDS	EST.					
SOURCE	Homo sapiens (human)					
ORGANISM	Homo sapiens					
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.					
REFERENCE	1 (bases 1 to 617)					
AUTHORS	Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R., Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and Kim,Y.S.					
TITLE	21C Frontier Korean EST Project 2001					
JOURNAL	Unpublished					
COMMENT	Contact: Kim YS Genome Research Center Korea Research Institute of Bioscience & Biotechnology 52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea Tel: +82-42-860-4470 Fax: +82-42-860-4409 Email: yongsung@mail.kribb.re.kr Plate: 27 row: H column: 07 High quality sequence stop: 617.					

FEATURES	SOURCE
1. <i>Age</i>	1. <i>Age</i>
2. <i>Gender</i>	2. <i>Gender</i>
3. <i>Marital Status</i>	3. <i>Marital Status</i>
4. <i>Education</i>	4. <i>Education</i>
5. <i>Occupation</i>	5. <i>Occupation</i>
6. <i>Income</i>	6. <i>Income</i>
7. <i>Religion</i>	7. <i>Religion</i>
8. <i>Political Affiliation</i>	8. <i>Political Affiliation</i>
9. <i>Health Status</i>	9. <i>Health Status</i>
10. <i>Travel History</i>	10. <i>Travel History</i>
11. <i>Employment Status</i>	11. <i>Employment Status</i>
12. <i>Family Size</i>	12. <i>Family Size</i>
13. <i>Home Ownership</i>	13. <i>Home Ownership</i>
14. <i>Vehicle Ownership</i>	14. <i>Vehicle Ownership</i>
15. <i>Insurance Coverage</i>	15. <i>Insurance Coverage</i>
16. <i>Subscription Services</i>	16. <i>Subscription Services</i>
17. <i>Charitable Contributions</i>	17. <i>Charitable Contributions</i>
18. <i>Volunteer Work</i>	18. <i>Volunteer Work</i>
19. <i>Real Estate Transactions</i>	19. <i>Real Estate Transactions</i>
20. <i>Financial Investments</i>	20. <i>Financial Investments</i>
21. <i>Medical History</i>	21. <i>Medical History</i>
22. <i>Legal Records</i>	22. <i>Legal Records</i>
23. <i>Criminal Record</i>	23. <i>Criminal Record</i>
24. <i>Academic Achievements</i>	24. <i>Academic Achievements</i>
25. <i>Professional Certifications</i>	25. <i>Professional Certifications</i>
26. <i>Travel Itineraries</i>	26. <i>Travel Itineraries</i>
27. <i>Hotel Reservations</i>	27. <i>Hotel Reservations</i>
28. <i>Airline Tickets</i>	28. <i>Airline Tickets</i>
29. <i>Rental Agreements</i>	29. <i>Rental Agreements</i>
30. <i>Utility Bills</i>	30. <i>Utility Bills</i>
31. <i>Bank Statements</i>	31. <i>Bank Statements</i>
32. <i>Credit Reports</i>	32. <i>Credit Reports</i>
33. <i>Tax Returns</i>	33. <i>Tax Returns</i>
34. <i>Insurance Claims</i>	34. <i>Insurance Claims</i>
35. <i>Medical Claims</i>	35. <i>Medical Claims</i>
36. <i>Legal Filings</i>	36. <i>Legal Filings</i>
37. <i>Academic Transcripts</i>	37. <i>Academic Transcripts</i>
38. <i>Professional Licenses</i>	38. <i>Professional Licenses</i>
39. <i>Travel Vouchers</i>	39. <i>Travel Vouchers</i>
40. <i>Hotel Receipts</i>	40. <i>Hotel Receipts</i>
41. <i>Airline Itineraries</i>	41. <i>Airline Itineraries</i>
42. <i>Rental Agreements</i>	42. <i>Rental Agreements</i>
43. <i>Utility Bills</i>	43. <i>Utility Bills</i>
44. <i>Bank Statements</i>	44. <i>Bank Statements</i>
45. <i>Credit Reports</i>	45. <i>Credit Reports</i>
46. <i>Tax Returns</i>	46. <i>Tax Returns</i>
47. <i>Insurance Claims</i>	47. <i>Insurance Claims</i>
48. <i>Medical Claims</i>	48. <i>Medical Claims</i>
49. <i>Legal Filings</i>	49. <i>Legal Filings</i>
50. <i>Academic Transcripts</i>	50. <i>Academic Transcripts</i>
51. <i>Professional Licenses</i>	51. <i>Professional Licenses</i>
52. <i>Travel Vouchers</i>	52. <i>Travel Vouchers</i>
53. <i>Hotel Receipts</i>	53. <i>Hotel Receipts</i>
54. <i>Airline Itineraries</i>	54. <i>Airline Itineraries</i>
55. <i>Rental Agreements</i>	55. <i>Rental Agreements</i>
56. <i>Utility Bills</i>	56. <i>Utility Bills</i>
57. <i>Bank Statements</i>	57. <i>Bank Statements</i>
58. <i>Credit Reports</i>	58. <i>Credit Reports</i>
59. <i>Tax Returns</i>	59. <i>Tax Returns</i>
60. <i>Insurance Claims</i>	60. <i>Insurance Claims</i>
61. <i>Medical Claims</i>	61. <i>Medical Claims</i>
62. <i>Legal Filings</i>	62. <i>Legal Filings</i>
63. <i>Academic Transcripts</i>	63. <i>Academic Transcripts</i>
64. <i>Professional Licenses</i>	64. <i>Professional Licenses</i>
65. <i>Travel Vouchers</i>	65. <i>Travel Vouchers</i>
66. <i>Hotel Receipts</i>	66. <i>Hotel Receipts</i>
67. <i>Airline Itineraries</i>	67. <i>Airline Itineraries</i>
68. <i>Rental Agreements</i>	68. <i>Rental Agreements</i>
69. <i>Utility Bills</i>	69. <i>Utility Bills</i>
70. <i>Bank Statements</i>	70. <i>Bank Statements</i>
71. <i>Credit Reports</i>	71. <i>Credit Reports</i>
72. <i>Tax Returns</i>	72. <i>Tax Returns</i>
73. <i>Insurance Claims</i>	73. <i>Insurance Claims</i>
74. <i>Medical Claims</i>	74. <i>Medical Claims</i>
75. <i>Legal Filings</i>	75. <i>Legal Filings</i>
76. <i>Academic Transcripts</i>	76. <i>Academic Transcripts</i>
77. <i>Professional Licenses</i>	77. <i>Professional Licenses</i>
78. <i>Travel Vouchers</i>	78. <i>Travel Vouchers</i>
79. <i>Hotel Receipts</i>	79. <i>Hotel Receipts</i>
80. <i>Airline Itineraries</i>	80. <i>Airline Itineraries</i>
81. <i>Rental Agreements</i>	81. <i>Rental Agreements</i>
82. <i>Utility Bills</i>	82. <i>Utility Bills</i>
83. <i>Bank Statements</i>	83. <i>Bank Statements</i>
84. <i>Credit Reports</i>	84. <i>Credit Reports</i>
85. <i>Tax Returns</i>	85. <i>Tax Returns</i>
86. <i>Insurance Claims</i>	86. <i>Insurance Claims</i>
87. <i>Medical Claims</i>	87. <i>Medical Claims</i>
88. <i>Legal Filings</i>	88. <i>Legal Filings</i>
89. <i>Academic Transcripts</i>	89. <i>Academic Transcripts</i>
90. <i>Professional Licenses</i>	90. <i>Professional Licenses</i>
91. <i>Travel Vouchers</i>	91. <i>Travel Vouchers</i>
92. <i>Hotel Receipts</i>	92. <i>Hotel Receipts</i>
93. <i>Airline Itineraries</i>	93. <i>Airline Itineraries</i>
94. <i>Rental Agreements</i>	94. <i>Rental Agreements</i>
95. <i>Utility Bills</i>	95. <i>Utility Bills</i>
96. <i>Bank Statements</i>	96. <i>Bank Statements</i>
97. <i>Credit Reports</i>	97. <i>Credit Reports</i>
98. <i>Tax Returns</i>	98. <i>Tax Returns</i>
99. <i>Insurance Claims</i>	99. <i>Insurance Claims</i>
100. <i>Medical Claims</i>	100. <i>Medical Claims</i>
101. <i>Legal Filings</i>	101. <i>Legal Filings</i>
102.	

```

1. .617
/organism="Homo sapiens"
/mol_type="rRNA"
/db_xref="taxon:9606"
/clone="S21SNU520-27-H07"
/sex="F"
/tissue_type="Stomach"
/cell_type="Floating aggregates"
/cell_line="SNU-520"
/lab_host="Top10F"
/clone_lib="S21SNU520"
/notes="Organ: Stomach; Vector: pTZ18RP1; Site_1: EcoRI;
Site 2: NotI; The poly (A) + RNA was dephosphorylated with
bacterial alkaline phosphatase (BAP) and then decapped
with tabacco acid pyrophosphatase (TAP). The decapped
intact mRNA was ligated with DNA-RNA linker including EcoR
I site by treatment of T4 RNA ligase and the first strand
cDNA was synthesized from oligo dT-selected mRNA by
priming with dT-tailed vector. The dT-tailed vector was
adjusted to have about 60nt. The cDNA vector was
circularized with E. coli DNA ligase after digestion of
EcoRI which site is also included in vector. An RNA strand
converted to a DNA strand by Okayama-Berg method. The
obtained cDNA vectors were used for transfection of
competent cells E. coli Top10F' by electroporation method.
The cDNA libraries constructed by this method are
full-length enriched cDNA library."
135 a 1274 C 182 G 124 T 2 others

```

BASE C
ORIGIN

Alignment Scores:	8.6e-104	Length:	617
Pred. No.:	1054.50	Matches:	199
Score:	97.56%	Conservative:	4
Percent Similarity:	97.07%	Mismatches:	1
Best Local Similarity:	45.03%	Indels:	1
Query Match:	12	Gaps:	1
DB:			

US-09-607-745-2 (1-435) x BM795149 (1-617)

76 AspGlyGluLeuAspCysProLeuGlyGluAspGluGluHisCysValLvsserPhePro 95

```

o 2 GACGGAGAGCTGCACTCTCCCTTTGGGGAGGACGAGGAGCACTGTGTCAAGAGCTTCCCC 61
Y 96 GluGlyProAlaValAlaValArgLeuSerLysAspArgSerThrLeuGlnValLeuAsp 115
o 62 GAAGGGCTCGAGTGGGAGTCGGCTCTCAAGGACCGATCCACACATGCAAGTGTGGAC 121
Y 116 SerAlaThrGlyAsnTrpPheSerAlaCysPheAspAsnPheThrGluAlaLeuAlaGlu 135
o 122 TCGGCCACAGGAACTGGTCTCTGCGCTGTGTGACAACTTCACAGAACTCTGCGCTGAG 181
Y 136 ThrAlaCysArgGlnMetGlyTyrSerSerLysProThrPheArgAlaValGluIleGly 155
o 182 ACAGCCTGTAGGAGATGGGCTACAGACGACAAACCCACTTTCAGAGCTGTGGAGTTGGC 241
Y 156 ProAspGlnAspLeuAspValValGluIlePheGluAsnSerGlnGluLeuArgMetArg 175
b 242 CCAGACCAAGATCTGGATGTGTGTAATACAGAAACAGCCAGGAGCTTCGATCGGG 301
Y 176 AsnSerSerGlyProCysLeuSerLysSerLeuValSerLeuHisCysLeuAlaCysGly 195
b 302 AACTCAAGTGGGCCCTGTCTCTCAGGCTCCCTGTCTCCCTGCACCTGTCTTGGCTGGGG 361
Y 196 LysSerLeuLysThrProArgValValGly---GlyGluGluAlaSerValAspSerTrp 214
b 362 AAGAGCCTGAGACCCCGCTGTGTGTGGNGTGGGGAAGGAGGCCCTCTGTGGATCTTGG 421
Y 215 ProTrpGlnValSerIleGlnTyrAspLysGlnHisValCysGlyGlySerIleLeuAsp 234
b 422 CTTTGGCAGGTACAGATACAGTACAGAAACAGACGCTGTGTGGAGGAGATCCTCGAC 481
Y 235 ProHisTrpValLeuThrAlaAlaHisCysPheArgLysHisThrAspValPheAsnTrp 254
b 482 CCCCACTGGGTCCTCAGCGACGCCACTGCTTCAGGAAACATACCGATGTGTTCAACTGG 541
Y 255 LysValArgAlaGlySerAspLysLeuGlySerPhePheProSerLeuAlaValAlaLysIle 274
b 542 AAGTGGGGGAGGAGCTCAACAAATGGGAGCTTCCCATCCCTGGCTGTGGCCAGATC 601
Y 275 IleIleIleGluPhe 279
b 602 ATCATCATTTGATTC 616

```

```

RESULT 8
US23218
OCUS
EFINITION
AGENCOURT 10154487 NCI CGAP Co24 Mus musculus cdna clone
IMAGE:6529864 5', mRNA sequence.
CCESION
US23218
US23218.1 GI:22833656
EYWORDS
EST.
SOURCE
Mus musculus (house mouse)
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 956)
NIH-MGC http://mgs.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: The Cepko Laboratory
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L14M14130 row: k column: 16
High quality sequence stop: 706.
FEATURES
Location/Qualifiers
1..956
/organism="Mus musculus"
/mol_type="mRNA"

```

```

/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:6529864"
/lab_host="DH10B (T1 phage-resistant)"
/clone_lib="NCI_CGAP_Co24"
/Note="Organ: colon; Vector: pCMV-SPORT6; Site: 1: NotI;
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dr.
Average insert size 1.6 kb. Constructed by Life
Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT 214 a 275 c 280 g 187 t
ORIGIN
Alignment Scores:
Pred. No.: 3,87e-102 Length: 956
Score: 1042.00 Matches: 213
Percent Similarity: 81.25% Conservative: 34
Best Local Similarity: 70.07% Mismatches: 48
Query Match: 44.49% Indels: 9
DB: 13 Gaps: 4
US-09-607-745-2 (1-435) x BU523218 (1-956)

```

```

QY 96 GluGlyProAlaValAlaValArgLeuSerLysAspArgSerThrLeuGlnValLeuAsp 115
Db 46 GAAAGCCCGGAGTGGCAGTCCGGCTCTCCAGGACAGATCCACCTGCGAGTGTGGAT 105
QY 116 SerAlaThrGlyAsnTrpPheSerAlaCysPheAspAsnPheThrGluAlaLeuAlaGlu 135
Db 106 GCAGCCACAGGACCTGGGCTCAGTCTGTTTCGACAACTTCACAGAACACTGGCCAC 165
QY 136 ThrAlaCysArgGlnMetGlyTyrSerSerLysProThrPheArgAlaValGluIleGly 155
Db 166 ACAGCCTGCAGACAGATGGCTATGACAGCCGCCCTTTTCAGACAGTGGAGATCCGT 225
QY 156 ProAspGlnAspLeuAspValValGluIleThrGluAsnSerGlnGluLeuArgMetArg 175
Db 226 CCAGATCAGAACTCCCTGTGTCTCAAGTCACAGAAACAGCCAGGAACCTCAGGTGAC 285
QY 176 AsnSerSerGlyProCysLeuSerGlySerLeuValSerLeuHisCysLeuAlaCysGly 195
Db 286 AATGGAGACGATCTTCGCTCAGGCTCCCTGGTTCTTGGCTGCTGATGTGGA 345
QY 196 LysSerLeuLysThrProArgValValGlyGlyGluAlaSerValAspSerTrpPro 215
Db 346 AAGAGCCTGAAGACTCTCTGTGTGTGTGAGTGGGTGGAGCCCTGTGGATCTTGGCCG 405
QY 216 TrpGluValSerIleGlnTyrAspLysGlnHisValCysGlyGlySerIleLeuAspPro 235
Db 406 TGGCAGGTACAGATCCAGTACAAACAGCAGCATGTCTGTGTGGGAGCATCTCGGATCCC 465
QY 236 HisTrpValLeuThrAlaAlaHisCysPheArgLysHisThrAspValPheAsnTrpLys 255
Db 466 CACTGGATCTCAGACAGCCCACTGCTTCAGAACTATCTTGTGTGTCAGACTGGAAG 525
QY 256 ValArgAlaGlySerAspLysLeuGlySerPheProSerLeuAlaValAlaLysIleIle 275
Db 526 GTCAGGCGAGGCTCAAAACATACACTCTGGGTAACTCTCCATCTCTTGGCTGTGGCCAAAGATCTTC 585
QY 276 IleIleGluPheAsnProMetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGln 295
Db 586 ATCGCTGAACCCCAATCTCTGTATCCCAAGAGAGGACATTCCTCTGTAAAGCTGAC 645
QY 296 PheProLeuThrPheSerGlyThrValArgProIleCysLeuProPhePheAspGluGlu 315
Db 646 ATGCCACTACATCTCAGGCTCAGTCAGGCCCATCTGCTGCTGCTCTCTGTAGAGGTG 705
QY 316 LeuThrProAlaThrProLeuTrpIleIleGlyTrp-GlyPheThrLysGlnAsn-GlyC 335
Db 706 CTTGTCCCGCCACACACAGTCTGGGTCTATGGATGGGGGCTTTACAGAAACCCGAG 765
QY 335 LysMetSerAspIleLeuLeuGln-AlaSerVal-GlnValIleAspSerThrArgCys 354
Db 766 GAAGATGTCTGTGATCTACTCAAGGATCATGTCCTCAGGTTCATCGACACCGAGGTG 825

```

ORIGIN

```

/ 354 s-AsnAlaAspAlaTyr---GlnGlyGluValThrGluLysMetMetCysAlaGlyI 373
/ 826 GCAATGCAGAGGATGCTACCAAGGGAAGGAGGACCGCTGAAATGGTGGGGGACAGG 885
/ 373 lePro---GluGlyGlyValAspThr---CysGlnGlyAspSer---GlyGlyProLeuM 390
/ 886 TACCCACCAAGGAGGAGGACACCCCTGCCAGGGGGGCCAGTGGCGGGGCCCTTT 945
/ 390 et 390
/ 946 TG 947

RESULT 9
LOCUS 4686292
DEFINITION UI-E-CRO-adt-g-09-0-UI.r1 UI-E-CRO Homo sapiens cDNA clone
ACCESSION EM686292.1 GI:18999550
VERSION 1
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene
discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
PUBMED 8889548
COMMENT Contact: Soares, MB
Coordinated Laboratory for Computational Genomics
University of Iowa
375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: bento-soares@uiowa.edu
Tissue Procurement: Dr. Gregg Hageman
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research
Genetics (www.resgen.com).
Seq primer: M13 Reverse.
Location/Qualifiers
1..569
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="UI-E-CRO-adt-g-09-0-UI"
/tissue_type="eye anterior segment"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/clone_lib="UI-E-CRO"
/notes="Organ: eye; Vector: pT73-Pac (Pharmacia) with a
modified polylinker; Site 1: EcoR I; Site 2: Not I;
UI-E-CRO is a cDNA library containing the following
tissue(s): eye anterior segment. The library was
constructed according to Bonaldo, Lennon and Soares,
Genome Research, 6:791-806, 1996. First strand cDNA
synthesis was primed with an oligo-dT primer containing a
Not I site. Double stranded cDNA was ligated to an EcoR I
adaptor, digested with Not I, and cloned directionally
into pT73-Pac vector. The oligonucleotide used to prime
the synthesis of first-strand cDNA contains a library tag
sequence that is located between the Not I site and the
(dT)18 tail. The sequence tag for this library is
AATCGCGAT. This library was created for the program: Gene
Discovery in the Visual System, supported by National Eye
Institute (NEI)."
```

BASE COUNT

124 a 169 c 156 g 120 t

```

US-09-607-745-2 (1-435) x BM686292 (1-569)

QY 173 ArgMetArgAsnSerSerGlyProCysLeuSerGlySer-LeuValSerLeuHisCysLe 192
Db 4 CGAGGGCGGAACCTCAAGTGGGCCCTCTCTCAGGCTCCCTTGGTCTCCCTGCACTGTCT 63
QY 192 uAlaCysGlyLysSerLeuLysThrProArgValValGlyGlyGlyGluAlaSerValAs 212
Db 64 TGCTGTGGGAAGAC-CTGAGAGACCCCGGTGTGGTGGGGAGAGGAGGCTCTGTGGA 122
QY 212 pSerTrpProTrpGlnValSerIleGlnTrpAspLysGlnHisValCysGlyGlySerIl 232
Db 123 TTCTTGGCCTTGGCAGGTCCAGCATCCAGTACGACAAACAGCACGCTCTGTGGAGGAGCAT 182
QY 232 eLeuAspProHisTrpValLeuThrAlaAlaHisCysPheArgLysHisThrAspValPh 252
Db 183 CTTGGACCCCGCATGGTCTCTCAGGAGCCCTGCTTCCAGAAACATACCGATGTGT 242
QY 252 eAsnTrpLysValArgAlaGlySerAspLysLeuGlySerPheProSerLeuAlaValAl 272
Db 243 CAACTGGAGGTGGGGCAGGCTCAGACAACTGGGCAGCTTCCCATCCCTGGCTGTGGC 302
QY 272 aLysIleIleIleIleGluPheAsnProMetTrpProLysAspAsnAspIleAlaLeuMe 292
Db 303 CAAGATCATCATCATTCATCAATTCACCCCATGTACCCCAAGACAAATGACATCGCCCTCAT 362
QY 292 tLysLeuGlnPheProLeuThrPheSerGlyThrValArgProIleCysLeuProPhePh 312
Db 363 GAAAGCTGCAGTTCCTTCTCAGGCACAGTCAGGCCCATCTGTCTGCCCTTCTT 422
QY 312 eAspGluGluLeuThrProAlaThrProLeuTrpIleIleGlyTrpGlyPheThrLysGl 332
Db 423 TGTATGAGGAGCTCACTCCAGCCACCCCATCTTGGATCATTTGGATGGGCTTTAGCAGCA 482
QY 332 nAsnGlyGlyLysMetSerAspIleLeuGlnAlaSerValGlnValIleAspSerTh 352
Db 483 GAATGGAGGGAAGATGTCTGACATCTGTCAGGGCGTCACTCAGTCAGTCCAGTCTTGCAGCAC 542
QY 352 rArgCysAsnAlaAspAlaTyr 360
Db 543 ACGGTGCAATGCAGACGATCGGTAC 567

RESULT 10
LOCUS BG822914
DEFINITION 602727917F1 NIH MGC_15 Homo sapiens cDNA clone IMAGE:4867108 5',
mRNA sequence.
ACCESSION BG822914
VERSION BG822914.1 GI:14170501
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 804)
AUTHORS NTH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-rc@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
```

DNA Sequencing by: NIH Intramural Sequencing Center
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: LLCMI735 row: b column: 05
 High quality sequence stop: 793.
 Location/Qualifiers

FEATURES

source
 1..804
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:4867108"
 /tissue_type="adenocarcinoma cell line"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="NIH_MGC_15"
 /notes="Organ: colon; Vector: pOTB7; Site 1: XhoI; Site 2:
 EcoRI; cDNA made by oligo-dT priming. Directionally
 cloned into EcoRI/XhoI sites using the following 5'
 adaptor: GGACGAG(G). Size-selected >500bp for average
 insert size 1.8kb. Library constructed by Ling Hong in
 the laboratory of Gerald M. Rubin (University of
 California, Berkeley) using ZAP-cDNA synthesis kit
 (Stratagene) and Superscript II RT (Life Technologies)"
 ASE COUNT 177 a 235 c 235 g 157 t
 RIGIN

Alignment Scores:

red. No.: 1.93e-92 Length: 804
 core: 951.50 Matches: 193
 Percent Similarity: 97.98% Conservativeness: 1
 Best Local Similarity: 97.47% Mismatches: 3
 Query Match: 40.63% Indels: 3
 Gaps: 12

S-09-607-745-2 (1-435) x BG822914 (1-804)

Y 1 MetAspProAspSerAspGlnProLeuAsnSerLeuAspValLysProLeuArgLysPro 20
 C 215 ATGGATCCTGACAGTATCAACCTCTGAACAGCTCGATGTCACAAACCCCTGGCAACCC 274
 Y 21 ArgTleProMetGluThrPheArgLysValGlyLeuProIleIleAlaLeuLeuSer 40
 C 275 CGATCCCCATGGAGACCTTCAGAAAGTGGGGATCCCATCATCATAGCACTACTGAGC 334
 Y 41 LeuAlaSerIleIleValValValLeuIleLysValIleLeuAspLysTyrTyrPhe 60
 b 335 CTGGCGAGTATCATCATTTGTGTTGCTCATCAAGGTGATTCCTGGATAAATACTACTTC 394
 Y 61 LeuCysGlyGlnProLeuHisPheIleProArgLysGlnLeuCysAspGlyGluLeuAsp 80
 b 395 CTCTGCGGGCAGCCTCTCCACTTCATCCGAGGAGCAGCTGTGTGACGGAGAGCTGGAC 454
 Y 81 CysProLeuGlyGluAspGluGluHisCysValLysSerPheProGluGlyProAlaVal 100
 b 455 TGTCCCTTGGGGAGGAGCAGAGACATGTGTCAAGAGCTTCCCGAGAGCCCTGCGAGT 514
 Y 101 AlaValArgLeuSerLysAspArgSerThrLeuGlnValLeuAspSerAlaThrGlyAsn 120
 b 515 GCAGTCCGCTCTCCAGGACCGATCCACATCGCAGGTGCTGGACTCGGCCACAGGGAAC 574
 Y 121 TrpPheSerAlaCysPheAspAsnPheThrGlu-AlaLeuAlaGluThrAlaCysArgG 140
 b 575 TGGTTCTCTGCTGCTTTTCACAACTTCACAGAAAGCTCTCGCTGAGACAGCTGTAGGCA 634
 Y 140 mMetGlyTyrSerSerLysProThrPheArgAlaValGluIleGlyProAspGlnAspLe 160
 b 635 GATGGGCTACAGACCAACCCACTTTTCAGAGCTGTGAGATTGGCCACAGACGAGTCT 694
 Y 160 uAspValValGluIleThrGluAsnSerGlnGluLeuArgMetArgAsnSerSerGlyP 180
 b 695 GGATGTCGTTGAAATACAGAAACAGCAGGAGCTTCGATGCGGAATCGGAACTCAAGTGGCC 754
 Y 180 oCysLeuSerGlySerLeuValSerLeuHisCysLeuAlaCysGlyLysSer 197

Db

RESULT 11

BG386903

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

755 -TGCTCTCAGGCTC-CTGGTCTCCTGAAGTGTCTTGCTGGGGAAGAGC 804

BG386903 1007 bp mRNA linear EST 12-MAR-2001
 602454702F1 NIH_MGC_15 Homo sapiens cDNA clone IMAGE:4582969 5',
 mRNA sequence.

BG386903

BG386903.1 GI:13280452

EST.

Homo sapiens (human)

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

NIH-MGC <http://mgs.nci.nih.gov/>

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: Ling Hong/Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: NIH Intramural Sequencing Center

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: LLCMI306 row: c column: 02

High quality sequence stop: 763.

Location/Qualifiers

1..1007

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:4582969"

/tissue_type="adenocarcinoma cell line"

/lab_host="DH10B (phage-resistant)"

/clone_lib="NIH_MGC_15"

/notes="Organ: colon; Vector: pOTB7; Site 1: XhoI; Site 2:

EcoRI; cDNA made by oligo-dT priming. Directionally

cloned into EcoRI/XhoI sites using the following 5'

adaptor: GGACGAG(G). Size-selected >500bp for average

insert size 1.8kb. Library constructed by Ling Hong in

the laboratory of Gerald M. Rubin (University of

California, Berkeley) using ZAP-cDNA synthesis kit

(Stratagene) and Superscript II RT (Life Technologies)"

BASE COUNT 240 a 291 c 293 g 183 t

ORIGIN

Alignment Scores:

Pred. No.: 4.53e-92 Length: 1007
 Score: 949.50 Matches: 207
 Percent Similarity: 84.17% Conservativeness: 11
 Best Local Similarity: 79.92% Mismatches: 22
 Query Match: 40.54% Indels: 19
 DB: 10 Gaps: 2

US-09-607-745-2 (1-435) x BG386903 (1-1007)

Qy 1 MetAspProAspSerAspGlnProLeuAsnSerLeuAspValLysProLeuArgLysPro 20

Db 215 ATGGATCCTGACAGTATCAACCTCTGAACAGCTCGATGTCACAAACCCCTGGCAACCC 274

Qy 21 ArgTleProMetGluThrPheArgLysValGlyLeuProIleIleAlaLeuLeuSer 40

Db 275 CGATCCCCATGGAGACCTTCAGAAAGTGGGGATCCCATCATCATAGCACTACTGAGC 334

Qy 41 LeuAlaSerIleIleValValValLeuIleLysValIleLeuAspLysTyrTyrPhe 60

Db 335 CTGGCGAGTATCATCATTTGTGTTGCTCATCAAGGTGATTCCTGGATAAATACTACTTC 394

Qy 61 LeuCysGlyGlnProLeuHisPheIleProArgLysGlnLeuCysAspGlyGluLeuAsp 80

```

b 395 CTCTGCGGCGAGCCTCTCCATCTCCGAGGAGAGCTGTGTGACGAGAGCTGCAC 454
y 81 CysProLeuGlyGluAspGluGluHisCysVallySerPheProGluGlyProAlaVal 100
b 455 TGTCCCTTGGGGAGGAGGAGGACATGTGTCAAGAGCTTCCCGAAGGCGCTGCGAGTG 514
y 101 AlaValArgLeuSerLysAspArgSerThrLeuGlnValLeuAspSerAlaThrGlyAsn 120
b 515 GCAGTCCGCTCTCCAGGACCGATCCACACTGCAGGTGTGGACTCGGCCACAGGAGAC 574
y 121 TrpPheSerAlaCysPheAspAspPheThrGluAlaLeuAlaGluThrAlaCysArgGln 140
b 575 TGGTTCTCTGCTGCTTTTCGACAACTTCACAGAGCTCTCGCTGAGACAGCCTGTAGGAC 634
y 141 MetGlyTyrSerSerLysProThrPheArgAlaValGluLeuGlyProAspGlnAspLeu 160
b 635 ATGGGCTTACAGCAACCAACCCACTTTCAGAGCTGTGGAGATTGGCCAGAACAGGATCTG 694
y 161 AspValValGluLeuThr-GluAsnSerGlnGluLeu-ArgMetArgAsnSerSer-Gly 179
b 695 GATGTCTGTGACATCACCAGAAACAGCCAGGAGCTTCGGAATCGGAACTCAAGTGGG 754
y 180 ProCysLeuSerGlySerLeuValSerLeuHisCysLeuAla-CysGlyLysSerLeu 199
b 755 CCTGGTCTCTCAGGAGTCCCTGCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 814
y 199 s-ThrProArgValValGlyGlyGluGluAlaSerVal---AspSerTrpProTrpGlnV 218
b 815 GAACCCCGTGTGCGCTTGTGGAAGAGGCCCCAGCGAGAACATCAAGCGCTTAGGGAAG 874
y 218 alser-lleGlnTyrAspLysGlnHisValCys----- 228
b 875 AGACAACTAGTACCAACACACCGAGTCTGTGAGAGAGACACTCGGGGAGCCCCACAG 934
y 229 -----GlyGlySerIleLeuAspProHisTrpValLeuThr 240
b 935 GGCACACTACAGGAAGACCTCGTATCAGAGCACTACCGTGTGTACA 979

ESTL12
2922656
CCUS
EFINITION
BO922656 920 bp mRNA linear EST 20-AUG-2002
IMAGE:6395909 5', mRNA sequence.
BO922656
BO922656.1 GI:22337687
EST.
Mus musculus (house mouse)
SOURCE
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Musinae; Mus.
1 (bases 1 to 920)
NITH-MGC http://mgi.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgaaps-remail.nih.gov
Tissue Procurement: The Cepko Laboratory
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: The I.M.A.G.E. Consortium (LMNL)
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L14M13891 row: f column: 06
High quality sequence stop: 599.
Location/Qualifiers
1. 920
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:6395909"

```

FEATURES
source

```

/lab_host="DH10B (T1 phage-resistant)"
/clone_lib="NCI CGAP Co24"
/notes="Ogan: colon; Vector: pCMW-SPORT6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.6 kb. Constructed by Life
Technologies. Note: this is a NCI CGAP library."
BASE COUNT 216 a 230 c 255 g 213 t 6 others
ORIGIN
Alignment Scores: 5.28e-86 Length: 920
Pred. No.: 893.00 Matches: 174
Percent Similarity: 75.95% Conservative: 25
Best Local Similarity: 66.41% Mismatches: 52
Query Match: 38.13% Indels: 12
DB: 13 Gaps: 2
US-09-607-745-2 (1-435) x BQ922656 (1-920)
Qy 181 CysLeuSerGlySerLeuValSerLeuHisCysLeuAlaCysGlyLysSerLeuLysThr 200
Db 9 TGCCTCTCAGGCTCCCTGGTTTCCTTGGCGCTGCTTGCCTGTGAAAGAGACCTGAAGACT 68
Qy 201 ProArgValValGlyGlyGluGluAlaSerValAspSerTrpProTrpGlnValSerIle 220
Db 69 CCTCGTGTGGTGGTGGGTTGGAGGCCCTGTGTGATTTTGGCGCTGGCAGGTCAGCATC 128
Qy 221 GlnTyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThr 240
Db 129 CAGTACACAGAGCAGCATGTCGTGTGGGGAGCATCTCTGGATCCCACTGGATCCTCACA 188
Qy 241 AlaAlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySer 260
Db 189 GCAGCCCACTGCTTCAGGAAGTATCTTGATGTGTCAAGCTGTGAAGTGTGAGGAGCTCA 248
Qy 261 AspLysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleGluPheAsn 280
Db 249 AACATATGGGTAACTCTCCATCTTGCCTGTGGCCCAAGATCTTCATCGCTGAACCCAT 308
Qy 281 ProMetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPhe 300
Db 309 CCTCTGTACCCCAAGAGAGAGCATGTCCCTTGTAAAGTGTGAGATGCCACTCACATTC 368
Qy 301 SerGlyThrValArgProIleCysLeuProPhePheAspGluGluLeuThrProAlaThr 320
Db 369 TCAGGCTCAGTACGGCCATCTGCTGCTTCTCTGATGAGGTGCTTGTCCAGCCACA 428
Qy 321 ProLeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyLysMetSerAspIle 340
Db 429 CCAGTCTGGGTCTATTGGATGGGCTTTACAGAAAGAAACCGAGAGAAAGATGTCTGACATG 488
Qy 341 LeuLeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAspAlaTyr 360
Db 489 CTACTGCAGCATCAGTCCAGGTCATCGACAGCACACGGTGCATCGCAGAGGATGCCTAC 548
Qy 361 GlnGlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThr 380
Db 549 GAAGGGGAGTACGACCGCTGAGATGTGTGTGAGGTACCCACAGGCTGTGCGAGACACC 608
Qy 381 CysGln-----GlyAspSerGlyGlyProLeuMetTyrGlnSerAspGlnTrp 396
Db 609 TGCAGGTGGGTCTCCGAAAGACCGCAGGAGAGTCTGTGTGAAGAGGGAATCCAAAGTCTAG 668
Qy 397 HisValValGlyIleValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyVal 416
Db 669 AGAGATGTGAACCC-CTACAGGCTGGATGCCAATGGAGTTGTAGCCGCTCAGTTTGTG 727
Qy 417 Tyr-----ThrLysValSerAlaTyrLeuAsnTrpIleTyrAsn 429
Db 728 ATAAAGGAAGGCATAATATGCTGTGAGAAAGTTAAANNAGGAGNATGGTGGATAGGTGAT 787
Qy 430 ValTrp 431

```



```

b      788 AGTGG 793

RESULT 13
G966811
OCUS
DEFINITION
60284306F1 NCI_CGAP_Co24 Mus musculus cDNA clone IMAGE:4989048 5',
mRNA sequence.
CCESSION
BG966811
ESSION
BG966811.1 GI:14354448
EYWORDS
EST.
Mus musculus (house mouse)
ORCE
Mus musculus
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 689)
NTH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
cDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L14M11002 row: k column: 01
High quality sequence stop: 680.
FEATURES
Location/Qualifiers
1..689
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:4989048"
/lab_host="DH10B (T1 phage-resistant)"
/clone_lib="NCI_CGAP_Co24"
/notes="Organ: colon; Vector: pCMV-SPORT6; Site:1: NotI;
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.6 kb. Constructed by Life
Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT 150 a 189 c 188 g 152 t
RIGIN

Alignment Scores:
red. No.: 5,54e-86 Length: 689
core: 891.00 Matches: 165
percent Similarity: 90.00% Conservative: 24
est Local Similarity: 78.57% Mismatches: 21
very Match: 38.04% Indels: 2
B: 12 Gaps: 0

S-09-607-745-2 (1-435) x BG966811 (1-689)
Y 226 HisValCysGlyGlySerIleLeuAspProHisTrpValLeuThrAlaAlaHisCysPhe 245
b 2 CATGTCTGTGTGGAGCATCTCTGGATCCCATCGTGGATCTCTCACAGCAGCCACTGCTTC 61
Y 246 ArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySerAspLysLeuGlySer 265
b 62 AGGAGATATCTTGATGTCTCAGCTGGAGGTGACGGCAGGCTCAACATCTGGGTAC 121
Y 266 PheProSerLeuAlaValAlaLysIleIleLeuGluPheAsnProMetTrpProLys 285
b 122 TCTCATCTCTGCTGTGGCAAGATCTTCATCGTGAACCCAACTCTGTACCCCAA 181
Y 286 AspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPheSerGlyThrValArg 305
b 182 GAGAAAGGACATGCTCTGTGTTAAGTGCAGATGCGACCTCACATCTCAGGTCAGT 241
Y 306 ProIleCysLeuProPhePheAspGluLeuThrProAlaThrProLeuTrpIleIle 325

```

```

Db      242 CCATCTGCCTGCCCTTCTCTGATGAGGTGTTGTCCCGACACACACAGTCTGGTCAIT 301
QY      326 GlyTrpGlyPheThrLysGlnAsnGlyGlyLysMetSerAspIleLeuLeuGlnAlaSer 345
Db      302 GGATGGGGCTTTACAGAGAAACGAGGAAAGATGCTCTGACATGCTACTGCAGGCATCA 361
QY      346 ValGlnValIleAspSerThrArgCysAsnAlaAspAlaValThrGlnGlyGluValThr 365
Db      362 GTCCAGGTCACTACACAGCAGCGGTCAATGACAGGATGCTCTACGAGGGGAAGTGACC 421
QY      366 GluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThrCysGlnGlyAspSer 385
Db      422 GGTGAGATGCTGTGTGTCAGGTACCCACAGGGTGGCAAGACACCTGCCA-GGTGACGT 480
QY      386 GlyGlyProLeuMetTrpGlnSerAspGlnTrpHisValValGlyIleValSerTrpGly 405
Db      481 GGTGGGGCCITTTGATGTACCATCTCTGACAAAGTGGCAGGTAGTAGCATCTGGGCGC 540
QY      406 TyrGlyCysGlyGlyProSerThrProGlyValThrLysValSerAlaValLeuAsn 425
Db      541 CATGATGCGCGG-CCAAGTACTCTCTGGAGTGTATACCAAGTCACTGCTATCTCAAC 599
QY      426 TrpIleTyrAsnValTrpLysAlaGluLeu 435
Db      600 TGGATCTACATGTTTCGGAAGTCTGAGATG 629

RESULT 14
BE616186
LOCUS
DEFINITION
601278758F1 NIH_MGC_39 Homo sapiens cDNA clone IMAGE:3610695 5',
mRNA sequence.
ACCESSION
BE616186
VERSION
BE616186.1 GI:9897785
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 879)
NTH-MGC http://mgi.nci.nih.gov/.
AUTHORS
National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL
Unpublished
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1CM267 row: c column: 16
High quality sequence stop: 733.
FEATURES
Location/Qualifiers
1..879
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:3610695"
/tissue_type="adenocarcinoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_39"
/notes="Organ: pancreas; Vector: pOTB7; Site:1: XhoI;
Site 2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Library constructed
by Ling Hong in the laboratory of Gerald M. Rubin
(University of California, Berkeley) using ZAP-cDNA
synthesis kit (Stratagene) and Superscript II RT (Life
Technologies)."
BASE COUNT 182 a 248 c 274 g 175 t
ORIGIN

```

Ligment Scores:

```

red. No.:          9.18e-86          Length:          879
core:             890.50          Matches:          198
Percent Similarity: 86.32%          Conservative:    4
est Local Similarity: 84.62%          Mismatches:    24
very Match:       38.02%          Indels:         11
B:                10              Gaps:           2

S-09-607-745-2 (1-435) x BE616186 (1-879)

Y 2 AspProAspSerAspGlnProLeuAsnSerLeuAspValLysProLeuArgLysProArg 21
b 198 GATCTGACAGTATCACTCTGACAGCTTCGATGTCACCAACCCCTCGGAAACCCCGT 257
Y 22 IleProMetGluThrPheArgLysValGlyLleProIlellelleAlaLeuLeuSerLeu 41
b 258 ATCCCATGAGACCTTCAGAAAGGTGGGGATCCCATCATCATGACACTACTGAGCGTG 317
Y 42 AlaSerllellelleValValValleuLeuLysVallelleLeuAspLysTyrPheLeu 61
b 318 GCGAGTATCATATTGTTGTTCTCATCAAGGTGATTCGTGATAAATACTACTTCCTC 377
Y 62 CysGlyGlnProLeuHisPheIleProArgLysGlnLeuCysAspGlyVleuLeuAspCys 81
b 378 TCGGGGAGCTCTCCATTTATCCCGAGGAGACAGCTGTGTGACGAGAGCTGACTGT 437
Y 82 ProLeuGlyGluAspGluGluHisCysValLysSerPheProGluGlyProAlaValAla 101
b 438 CCGTTGGGGAGGACGAGGACGACTGTGTCAAGAGCTTCCCGAAGGGCTCGAGTGCA 497
Y 102 ValArgLeuSerLysAspArgSerThrLeuGlnValLeuAspSerAlaThrGlyAsnTrp 121
b 498 GTCCGCTCTCCAGAGCCGATCCACATCGACGTGCTGCGACTCGGC-ACAGGGAACG 556
Y 122 PheSerAlaCysPheAspAsnPhenThrGluAlaLeuAlaGluThrAlaCysArgGlnMet 141
b 557 TTCTCTGCTGTTTCGACAACTTTCACAGAGCTTCGCTGAGACAGC-TGATGGCAGATG 615
Y 142 GlyTyrSerLysProThrPheArgAlaValGluLleGlyPro-AspGlnAspLeuAs 161
b 616 GGCTACAGCAGCAAAACACATTTTCAGAGCTGTGGAGATGGGCCAGGACAGGACTGGA 675
Y 161 pValValGluLleThrGluAsnSerGlnGluLeu-ArgMetArgAsnSerSerGlyProc 181
b 676 TGTGTTGTAATCATCAGAACACAGCAGGAGCTTTCGCATCGCAACTCAAGTGGGCC-T 734
Y 181 YsLeuSerGlySerLeuVal-SerLeuHisCysLeuAlaCysGlyLysSerLeuLysThr 200
b 735 GTCTCTCAGGGTCCCTGCTCTTCCTCGACACTGTTGCTGGGGAAG---CTGAAACCC 791
Y 201 ProArgValValGlyGlyGluGluAlaSerValAspSerTrpProTrpGlnValSerIle 220
b 792 CCCCTGTTGGGGGGGGGGGGGGGCC-----TGGGGGATATATGCGGTGGGG 839
Y 221 GlnTyrAspLysGlnHisValCysGlyClySerIle 232
b 840 GGACATCCCGCCACAGCGCGCTGGGGGGGTCTGTG 875

```

ESULT 15

```

B142902
OCUS
EFINITION K-BST0196799 L11SN0354s1 Homo sapiens cDNA clone L11SN0354e1-20-D07
5', mRNA sequence.
CCSSION
ERSON CB142902.1 GI:28119586
EYWORDS EST.
OURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 515)
Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.N., Park,H.S., Kim,S. and

```

Kim, Y.S.
21C Frontier Korean EST Project 2001
Unpublished
Contact: Kim YS
Genome Research Center
Korea Research Institute of Bioscience & Biotechnology
52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea
Tel: +82-42-860-4470
Fax: +82-42-860-4409
Email: yongsung@mail.kribb.re.kr
Plate: 20 row: D column: 07
High quality sequence stop: 515.
Location/Qualifiers
1. 515

TITLE
JOURNAL
COMMENT

FEATURES

source

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="L11SN0354s1-20-D07"
/sex="M"
/tissue_type="Liver"
/cell_type="Polygonal"
/cell_line="SNU-354"
/lab_host="Top10P"
/clone_lib="L11SN0354s1"
/note="Organ: Liver; Vector: pCNS-D2; Site 1: EcoRI;
Site 2: NotI; The poly (A)+ RNA was dephosphorylated with
bacterial alkaline phosphatase (BAP) and then decapped
with tabacco acid pyrophosphatase (TAP). The decapped
intact mRNA was ligated with DNA-RNA linker including
EcoRI site by treatment of T4 RNA ligase and the first
strand cDNA was synthesized from oligo dt-selected mRNA by
priming with dt-tailed vector. The dt-tailed vector was
adjusted to have about 60nt. The cDNA vector was
circularized with E. coli DNA ligase after digestion of
EcoRI which site is also included in vector. An RNA strand
converted to a DNA strand by Okayama-Berg method. The
obtained cDNA vectors were used for transformation of
competent cells E. coli Top10P' by electroporation method.
The cDNA libraries constructed by this method are
full-length enriched cDNA library. After analyzing and
sequencing about 2,000 - 3,000 colonies in original cDNA
library, the abundant cDNAs were selected and amplified by
PCR reaction using vector region primer including T7
promotor as 5' primer and N(dt)14 as 3' primer. The PCR
products were used as template for synthesis of
biotinylated single stranded RNA by in vitro transcription
reaction. The synthesized RNA probes were hybridized with
antisense single stranded cDNAs prepared from original
library and incubated with avidin-gel. After removing
DNA-RNA hybrids by centrifuge, the subtracted cDNA
libraries were constructed by transfection of the
remaining DNA into competent cells E. coli Top10P' with
electroporation method."

BASE COUNT 116 a 139 c 150 g 110 t
ORIGIN

Alignment Scores:

Pred. No.: 1.57e-85 Length: 515
Score: 885.00 Matches: 160
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 37.79% Indels: 0
DB: 14 Gaps: 0

US-09-607-745-2 (1-435) x CB142902 (1-515)

Qy 276 llelleGluPheAsnProMetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGln 295
Db 2 ATCATTGAATTCACCCCATGTATCCCAAGACATGACATCGCCCTCATGAGGTCAG 61
Qy 296 PheProLeuThrPheSerGlyThrValArgProIleCysLeuProPhePheAspGluGlu 315
Db 62 TTCCCACTACTTTCTCAGGCACAGTCAGGCCCATCTGTCTGCCCTTCTTTGATGAGGAG 121

```
y      316 LeuThrProAlaThrProLeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyGly 335
b      122 CTCACTCCAGCCACCCCACTCTGGATCATTTGGATGGGGCTTTACGAAGCAGAAATGGAGGG 181

y      336 LysMetSerAspIleLeuLeuGlnAlaSerValGlnValIleAspSerThrArgCysAsn 355
b      182 AAGATGTCCTGACATACCTGTCAGGCGTCAGTCCAGGTCAATTGACAGCACACGGTGCAT 241

y      356 AlaAspAspAlaTyrGlnGlnGlyValThrGluLysMetMetCysAlaGlyIleProGlu 375
b      242 GCAGACGATGCCGTACAGGGGAGTCACCGAGAAGATGATGTGCAGGATCCCGGAA 301

y      376 GlyGlyValAspThrCysGlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGln 395
b      302 GGGGGTGTGGACACCTGCCAGGGTGACAGTGGTGGGCCCTGATGTACCAATCTGACCAG 361

y      396 TrpHisValValGlyIleValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGly 415
b      362 TGGCATGTGTGGGCATCGTTAGTTGGGGCTATGGCTGCGGGGGCCCGAGCACCCACGGA 421

y      416 ValTyrThrLysValSerAlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu 435
b      422 GTATACACCAAGGTCTCAGCCTATCTCACTGGAICTACATGTCCTGGAGGCTGAGCTG 481
```

earch completed: November 23, 2003, 12:05:54
ob time : 3113.27 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

protein - protein search, using sw model

on: November 23, 2003, 07:38:25 ; Search time 42.9766 Seconds
(without alignments)
1078.451 Million cell updates/sec

icle: US-09-607-745-9

sequence: 1 MDSKSGSQSRLLLLLVSN.....LNWYVWKAELSRHHHHH 292

oring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

sarched: 1107863 seqs, 158726573 residues

otal number of hits satisfying chosen parameters: 1107863

inimum DB seq length: 0

aximum DB seq length: 2000000000

at-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

atabase :

A.Geneseq 19Jun03.*
1: /SIDSL1/gcgdata/geneseq/geneseq-embl/AA1980.DAT.*
2: /SIDSL1/gcgdata/geneseq/geneseq-embl/AA1981.DAT.*
3: /SIDSL1/gcgdata/geneseq/geneseq-embl/AA1982.DAT.*
4: /SIDSL1/gcgdata/geneseq/geneseq-embl/AA1983.DAT.*
5: /SIDSL1/gcgdata/geneseq/geneseq-embl/AA1984.DAT.*
6: /SIDSL1/gcgdata/geneseq/geneseq-embl/AA1985.DAT.*
7: /SIDSL1/gcgdata/geneseq/geneseq-embl/AA1986.DAT.*
8: /SIDSL1/gcgdata/geneseq/geneseq-embl/AA1987.DAT.*
9: /SIDSL1/gcgdata/geneseq/geneseq-embl/AA1988.DAT.*
10: /SIDSL1/gcgdata/geneseq/geneseq-embl/AA1989.DAT.*
11: /SIDSL1/gcgdata/geneseq/geneseq-embl/AA1990.DAT.*
12: /SIDSL1/gcgdata/geneseq/geneseq-embl/AA1991.DAT.*
13: /SIDSL1/gcgdata/geneseq/geneseq-embl/AA1992.DAT.*
14: /SIDSL1/gcgdata/geneseq/geneseq-embl/AA1993.DAT.*
15: /SIDSL1/gcgdata/geneseq/geneseq-embl/AA1994.DAT.*
16: /SIDSL1/gcgdata/geneseq/geneseq-embl/AA1995.DAT.*
17: /SIDSL1/gcgdata/geneseq/geneseq-embl/AA1996.DAT.*
18: /SIDSL1/gcgdata/geneseq/geneseq-embl/AA1997.DAT.*
19: /SIDSL1/gcgdata/geneseq/geneseq-embl/AA1998.DAT.*
20: /SIDSL1/gcgdata/geneseq/geneseq-embl/AA1999.DAT.*
21: /SIDSL1/gcgdata/geneseq/geneseq-embl/AA2000.DAT.*
22: /SIDSL1/gcgdata/geneseq/geneseq-embl/AA2001.DAT.*
23: /SIDSL1/gcgdata/geneseq/geneseq-embl/AA2002.DAT.*
24: /SIDSL1/gcgdata/geneseq/geneseq-embl/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

result No.	Score	Query Match	Length	ID	Description
1	1600	100.0	292	23	Protease D-G catay
2	1266	79.1	317	22	Human protein sequ
3	1266	79.1	405	24	Human expressed pr
4	1266	79.1	405	23	Human expressed pr
5	1266	79.1	406	23	CUA8 sequence. Ho
6	1266	79.1	423	22	Human transmembran
7	1266	79.1	423	24	Human colorectal c
8	1266	79.1	423	24	Human expressed pr
9	1266	79.1	423	24	Human expressed pr

10	1266	79.1	423	24	ABU049323	Human expressed pr
11	1266	79.1	432	21	AAV99417	Human PRO1570 (UNQ
12	1266	79.1	432	22	AAU29188	Human PRO polypept
13	1266	79.1	432	22	AAU29188	Human PRO1570. Ho
14	1266	79.1	432	22	AAU29188	Protein of the inv
15	1266	79.1	432	23	ABG95906	Human secreted/tra
16	1266	79.1	432	23	AAU76535	Tumour-associated
17	1266	79.1	432	24	ABU71276	Human PRO1570 prot
18	1266	79.1	432	24	ABU71561	Human secreted pol
19	1266	79.1	432	24	ABU72007	Novel human secret
20	1266	79.1	432	24	ABU72164	Human PRO polypept
21	1266	79.1	432	24	ABU65733	Human secreted/tra
22	1266	79.1	432	24	ABU66066	Novel human secret
23	1266	79.1	432	24	ABU67570	Human secreted/tra
24	1266	79.1	432	24	ABU65428	Human PRO polypept
25	1266	79.1	432	24	ABU58564	Human secreted/tra
26	1266	79.1	432	24	ABU56100	Human PRO polypept
27	1266	79.1	432	24	ABU57095	Human expressed pr
28	1266	79.1	432	24	ABU04921	Human expressed pr
29	1266	79.1	432	24	ABU04929	Human expressed pr
30	1266	79.1	432	24	ABU04934	Human expressed pr
31	1266	79.1	432	24	ABU10674	Human protease HUP
32	1266	79.1	435	20	AAV06437	Human seripancrin
33	1266	79.1	435	22	AAV72558	Human serine prote
34	1266	79.1	435	23	AAV78577	Human expressed pr
35	1266	79.1	435	24	ABU04912	Human expressed pr
36	1266	79.1	435	24	ABU04930	Human expressed pr
37	1266	79.1	435	24	ABU04931	Human expressed pr
38	1266	79.1	437	23	AAV79359	CJA8 preferred seq
39	1266	79.1	437	24	ABU04915	Human expressed pr
40	1266	79.1	437	24	ABU04920	Human expressed pr
41	1266	79.1	437	24	ABU04935	Human expressed pr
42	1266	79.1	437	24	ABU04936	Human expressed pr
43	1264	79.0	233	18	AAW22986	Human serine prote
44	1253	78.3	492	22	AAV72559	Human seripancrin
45	1253	78.3	492	24	ABU04913	Human expressed pr

ALIGNMENTS

RESULT 1

AAV78578

ID AAG78578 standard; Protein; 292 AA.

XX AAG78578;

DT 07-MAY-2002 (first entry)

XX Protease D-G cataytic domain fusion gene construct amino acid sequence.

XX Serine protease; D-G; human; zymogen; enzyme; cytostatic;

XX antiinflammatory; dermatological; anticoagulation; cancer;

XX skin disorder; neuropathic pain; inflammatory disorder;

XX coagulation diathesis; thrombosis; laundry detergent; skin care;

XX Gene therapy.

XX Homo sapiens.

XX Key Location/Qualifiers

FT Peptide

FT Protein

FT Domain

FT Misc-difference

FT Region

XX WO200202011-A1.

D 10-JAN-2002.
X
F 08-JUN-2001; 2001WO-US18568.
R 30-JUN-2000; 2000US-0607745.
X (ORTH) ORTHO-MCNEIL PHARM INC.
X Darrow AL, Qi J, Andrade-Gordon P;
I WPI; 2002-106601/14.
R N-PSDB; AA164290.
X
I Nucleic acid encoding a serine protease called D-G protein which is
I useful for identifying modulators that are useful for treating a
I condition which is mediated by protease D-G, e.g. cancer, skin
I disorders, or neuropathic pain -
X
S Claim 13; Fig 4B; Bipp; English.
X
C The invention relates to an isolated and purified nucleic acid that
C encodes a serine protease called D-G protein. The activity of the protein
C of the invention may be described as cytostatic, antiinflammatory,
C dermatological and anticoagulation. The serine protease of the invention
C is a member of the trypsin/chymotrypsin-like (S1) serine protease family,
C which play an important role in processes such as digestion and
C regulatory amplification cascades through the proteolytic activation of
C inactive zymogen precursors. Protease D-G modulating compounds are useful
C for treating a condition which is mediated by protease D-G, e.g. cancer,
C skin disorders, neuropathic pain, inflammatory disorders, or coagulation
C diathesis/thrombosis. The polynucleotide encoding the protease is useful
C for identifying, detecting or isolating mutant forms of DNA molecules
C encoding the protease. The protease is useful for identifying modulators
C of the functional protease. The D-G protein can be used for formulation
C of compositions for laundry detergents and skin care products. Protease
C D-G gene therapy may be used to introduce protease D-G into the cells of
C target organisms. As the D-G protein is derived from a human, it is less
C likely to produce an allergic reaction in sensitive individuals when used
C in formulations for laundry detergents and skin care products. The
C current sequence represents the protease D-G catalytic domain in the
C zymogen activation construct amino acid sequence.

X
Q Sequence 292 AA;

Query Match 100.0%; Score 1600; DB 23; Length 292;
Best Local Similarity 100.0%; Pred. No. 7.5e-144;
Matches 292; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 1 MDSKSSQKSRLLLLLVSNLLLCQGVSDYKDDDDVDAALAAAFDDDDKTVGGYALDV 60
b 1 MDSKSSQKSRLLLLLVSNLLLCQGVSDYKDDDDVDAALAAAFDDDDKTVGGYALDV 60
Y 61 DSWPQVSIQYDKQVCGGSIIDPHWVLTAAHCFRKHDTDFVKNVKGSDKLGSPSLAV 120
b 61 DSWPQVSIQYDKQVCGGSIIDPHWVLTAAHCFRKHDTDFVKNVKGSDKLGSPSLAV 120
Y 121 AKIIIEFNPMYPKNDIALMKLQPLFPFSGVTRPICLPFFDEELTPATPLIIGWGFTK 180
b 121 AKIIIEFNPMYPKNDIALMKLQPLFPFSGVTRPICLPFFDEELTPATPLIIGWGFTK 180
Y 181 QNGGRMSDILLQASVQVIDSTRCNADDAVQGEVTEKMCAGIPGEGVDTFCQDSSGGLMY 240
b 181 QNGGRMSDILLQASVQVIDSTRCNADDAVQGEVTEKMCAGIPGEGVDTFCQDSSGGLMY 240
Y 241 QSDQWHVGVISWGVCGGSPGTVYTKVSAVLMNTYNNWKALSRHHHHH 292
b 241 QSDQWHVGVISWGVCGGSPGTVYTKVSAVLMNTYNNWKALSRHHHHH 292

RESULT 2

AM25633

D AM25633 standard; Protein; 317 AA.

X

AC AAM25633;
XX 16-OCT-2001 (first entry)
XX Human protein sequence SEQ ID NO:1148.
XX
KW Human; cancer; ulcer; HIV infection; human immunodeficiency virus;
KW antiinflammatory; antirheumatic; antiarthritic; immunosuppressive;
KW antibacterial; endocrine; cardiac; central nervous system; virucide;
KW anti-HIV; fungicide; antimutagen; cardiovascular; antianaemic; anaemia;
KW antiaggregant; haemostatic; vulnery; antiulcer; osteopathic; eczema;
KW dermatological; antiallergic; antiasthmatic; antidiabetic; cytostatic;
KW neuroprotective; antidepressant; nootropic; antiparkinsonian; infection;
KW immunomulant; gene therapy; antisense therapy; vaccine; inflammation;
KW antianaphylactic; rheumatoid arthritis; septic shock; pancreatitis;
KW cardiac dysfunction; neuropathology; cardiac anaphylaxis; autoimmunity;
KW genetic disease; haematopoietic disorder; platelet disorder; asthma;
KW thrombocytopaenia; osteoporosis; diabetes; severe combined immunodeficiency;
KW allergic rhinitis; diabetes; multiple sclerosis; depression;
KW Alzheimer's disease; Parkinson's disease; neurodegenerative disorder;
KW neurological disorder.
XX Homo sapiens.
OS WO200153455-A2.
XX 26-JUL-2001.
XX 22-DEC-2000; 2000WO-US35017.
XX 23-DEC-1999; 99US-0471275.
PR 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
XX (HYSE-) HYSEQ INC.
XX Tang YT, Liu C, Drmanac RT;
PI WPI; 2001-457603/49.
XX N-PSDB; AAH99574.
XX Isolated human polynucleotides encoding polypeptides, useful for the
XX treatment and diagnosis of e.g. cancer, ulcers and HIV infection -
XX
XX Claim 20; Page 238; 1217pp; English.
XX
CC AAH99166 to AAH99904 encode the human proteins given in AAM25225 to
CC AAM25963. The proteins can have activities based on the tissues and
CC cells they are expressed in, such as: antiinflammatory; antirheumatic;
CC antiarthritic; immunosuppressive; antibacterial; endocrine; cardiac;
CC central nervous system; virucide; anti-HIV; fungicide; antimutagen;
CC cardiovascular; antianaemic; antiaggregant; haemostatic; vulnery;
CC antiulcer; osteopathic; dermatological; antiallergic; antiasthmatic;
CC antidiabetic; cytostatic; neuroprotective; antidepressant; nootropic;
CC antiparkinsonian; and immunostimulant. The proteins and polynucleotides
CC encoding them can be used in gene therapy, antisense therapy and vaccine
CC production. The proteins and polynucleotides are useful for screening for
CC agonists or antagonists of a protein and for the treatment and diagnosis
CC of disorders associated with the activity of a protein e.g. inflammation,
CC rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,
CC neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal
CC infections, autoimmunity, genetic diseases, haematopoietic disorders,
CC anaemia, platelet disorders, thrombocytopaenia, wounds, burns, ulcers,
CC osteoporosis, severe combined immunodeficiency, eczema, allergic
CC rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression,
CC Alzheimer's disease, Parkinson's disease, neurodegenerative and
CC neurological disorders.
XX
XX Sequence 317 AA;

Query Match 79.1%; Score 1266; DB 22; Length 317;
Best Local Similarity 97.4%; Pred. No. 4.9e-112;
Matches 228; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

```
Y 51 KIVGYALDVDSWPFQVSIQYDKQHVCGGSLDPHWLTAHCRKHTDVFVNWKVRAGSD 110
b 84 RVVGGEEASVDSWPFQVSIQYDKQHVCGGSLDPHWLTAHCRKHTDVFVNWKVRAGSD 143
Y 111 KLGSPFSLAVAKIIIEFNPMYPKNDIALMKLOPPLTFSGTVRPICLPFFDEBLTPATP 170
b 144 KLGSPFSLAVAKIIIEFNPMYPKNDIALMKLOPPLTFSGTVRPICLPFFDEBLTPATP 203
Y 171 LWIIGWGTQKQGGKMSDILLQASVQVIDSTRCNADDAYQGEVTEKMKCAGIPGGVDTC 230
b 204 LWIIGWGTQKQGGKMSDILLQASVQVIDSTRCNADDAYQGEVTEKMKCAGIPGGVDTC 263
Y 231 QGDSGGPLMYQSDQHVHVGVISWVGCGGPGSTPGVYTKVSAYLNIYVNWKAEL 284
b 264 QGDSGGPLMYQSDQHVHVGVISWVGCGGPGSTPGVYTKVSAYLNIYVNWKAEL 317

RESULT 3
BU04916
D ABU04916 standard; Protein; 405 AA.
X C
X C ABU04916;
X X
X X 29-JAN-2003 (first entry)
X X
X X Human expressed protein tag (EPT) #1582.
E X
X X Translational profiling; expressed protein tag; EPT; kinase;
W phosphatase; protease; protease inhibitor; transporter;
W cytoskeletal protein; receptor; transcription factor; cancer; MHC;
W major histocompatibility complex; myeloma; colon cancer;
W gastric cancer; adenocarcinoma; sarcoma; melanoma; lymphoma;
W leukaemia.
X X
X X Homo sapiens.
X X
X N WO200278524-A2.
X X
X D 10-OCT-2002.
X X
X P 28-MAR-2002; 2002WO-US09671.
X X
X R 28-MAR-2001; 2001US-279495P.
X R 21-MAY-2001; 2001US-292544P.
X R 08-AUG-2001; 2001US-310801P.
X R 01-OCT-2001; 2001US-326370P.
X R 04-DEC-2001; 2001US-336780P.
X R 20-FEB-2002; 2002US-358985P.
X X
X A (ZYCO-) ZYCOS INC.
X X
X X Chicx RM, Tomlinson AJ, Urban RG;
X I WPI; 2003-040607/03.
X X
X T New polypeptides (e.g. kinases, phosphatases, proteases, transporters,
T cytoskeletal proteins, receptors or transcription factors), useful for
T treating cancer, e.g. colon cancer, gastric cancer, sarcoma, lymphoma
T or leukemia -
X X
X S Example 2; SEQ ID No 1582; 134pp; English.
X X
X C The invention describes a purified polypeptide, which comprises a
X fragment of a kinase, phosphatase, protease, protease inhibitor,
X transporter, cytoskeletal protein, receptor or transcription factor.
X The polypeptide is useful as an immunogenic composition for eliciting
X in a mammal an immunogenic response directed against any of the purified
X polypeptide. The purified polypeptide, or the antibody that binds to
X this polypeptide, is useful for treating cancer. The polypeptide is
X also useful for identifying compounds that binds to a naturally
X processed class I or class II MHC-binding polypeptide. The polypeptides
X and polynucleotides are particularly useful for treating or preventing
```

```
CC myeloma, colon cancer, gastric cancer, adenocarcinoma, sarcoma, melanoma,
CC lymphoma or leukaemia. These are also useful for screening agents for
CC treating the above mentioned diseases. This sequence represents an
CC expressed protein tag (EPT) isolated from human tissue for translational
CC profiling.
CC Note: This sequence does not appear in the printed specification but was
CC obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
XQ Sequence 405 AA;
Query Match 79.1%; Score 1266; DB 24; Length 405;
Best Local Similarity 97.4%; Pred. No. 6.6e-112;
Matches 228; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
QY 51 KIVGYALDVDSWPFQVSIQYDKQHVCGGSLDPHWLTAHCRKHTDVFVNWKVRAGSD 110
Db 172 RVVGGEEASVDSWPFQVSIQYDKQHVCGGSLDPHWLTAHCRKHTDVFVNWKVRAGSD 231
QY 111 KLGSPFSLAVAKIIIEFNPMYPKNDIALMKLOPPLTFSGTVRPICLPFFDEBLTPATP 170
Db 232 KLGSPFSLAVAKIIIEFNPMYPKNDIALMKLOPPLTFSGTVRPICLPFFDEBLTPATP 291
QY 171 LWIIGWGTQKQGGKMSDILLQASVQVIDSTRCNADDAYQGEVTEKMKCAGIPGGVDTC 230
Db 292 LWIIGWGTQKQGGKMSDILLQASVQVIDSTRCNADDAYQGEVTEKMKCAGIPGGVDTC 351
QY 231 QGDSGGPLMYQSDQHVHVGVISWVGCGGPGSTPGVYTKVSAYLNIYVNWKAEL 284
Db 352 QGDSGGPLMYQSDQHVHVGVISWVGCGGPGSTPGVYTKVSAYLNIYVNWKAEL 405

RESULT 4
ABU04937
ID ABU04937 standard; Protein; 405 AA.
XX
X AC ABU04937;
XX
XX 29-JAN-2003 (first entry)
XX
XX Human expressed protein tag (EPT) #1603.
XX
XX Translational profiling; expressed protein tag; EPT; kinase;
KW phosphatase; protease; protease inhibitor; transporter;
KW cytoskeletal protein; receptor; transcription factor; cancer; MHC;
KW major histocompatibility complex; myeloma; colon cancer;
KW gastric cancer; adenocarcinoma; sarcoma; melanoma; lymphoma;
KW leukaemia.
XX
XX Homo sapiens.
XX
XX WO200278524-A2.
XX
XX 10-OCT-2002.
XX
XX 28-MAR-2002; 2002WO-US09671.
XX
XX 28-MAR-2001; 2001US-279495P.
PR 21-MAY-2001; 2001US-292544P.
PR 08-AUG-2001; 2001US-310801P.
PR 01-OCT-2001; 2001US-326370P.
PR 04-DEC-2001; 2001US-336780P.
PR 20-FEB-2002; 2002US-358985P.
XX
XX (ZYCO-) ZYCOS INC.
XX
XX Chicx RM, Tomlinson AJ, Urban RG;
XX WPI; 2003-040607/03.
XX
XX New polypeptides (e.g. kinases, phosphatases, proteases, transporters,
XX cytoskeletal proteins, receptors or transcription factors), useful for
XX treating cancer, e.g. colon cancer, gastric cancer, sarcoma, lymphoma
```


W MTSP; protease domain; neoplastic disease; tumour; cancer; cytostatic;
 W lung carcinoma; colon adenocarcinoma; ovarian carcinoma; gene therapy;
 W transmembrane protease serine 4; TMRSS4.

X Homo sapiens.

X X Key Location/Qualifiers
 H Region 190..194
 T /note= "Conserved motif"

X WO200157194-A2.

X 09-AUG-2001.

X 02-FEB-2001; 2001WO-US03471.

X 03-FEB-2000; 2000US-0179982.

X 18-FEB-2000; 2000US-0183542.

X 22-JUN-2000; 2000US-0213124.

X 26-JUL-2000; 2000US-0220970.

X 08-SEP-2000; 2000US-0657986.

X 22-SEP-2000; 2000US-0234840.

X (CORV-) CORVAS INT INC.

X Madison EL, Ong EO, Yeh J;

X WPI; 2001-488877/53.

X N-PSDB; AAD13169.

X Disclosure; Page 255-256; 256pp; English.

The invention relates to transmembrane serine proteases and their corresponding nucleotides and the protease domain of a type-II membrane-type serine protease (MTSP). MTSP is useful for identifying compounds that modulate or inhibit its proteolytic activity and for formulating a medicament for treating neoplastic disease. MTSP and its corresponding nucleotides are useful in preventing or treating tumours or cancers such as lung carcinoma, colon adenocarcinoma and ovarian carcinoma, in diagnostics and in hybridisation assays. MTSP is useful as a diagnostic marker for tumour development, growth and/or progression and as immunogens to generate antibodies that specifically bind to it. MTSP DNA is useful in a yeast two-hybrid system and in gene therapy. The present sequence is human transmembrane protease serine 4 (TMRSS4) protein.

X Sequence 423 AA;

Query Match 79.1%; Score 1266; DB 22; Length 423;
 Best Local Similarity 97.4%; Pred. No. 7e-112;
 Matches 228; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Y 51 KIYGVYALVDVSNPQVSIQYDKQHVCGGSIIDPHWVLTAAHCFRKHDTVFNNKVRAGSD 110

190 RYVGGSEASVDSNPQVSIQYDKQHVCGGSIIDPHWVLTAAHCFRKHDTVFNNKVRAGSD 249

111 KLGSPSLAVAKIIIEFNPMYKNDIALMKLQPLTFSGTVRPICLPFFDELTPTATP 170

250 KLGSPSLAVAKIIIEFNPMYKNDIALMKLQPLTFSGTVRPICLPFFDELTPTATP 309

171 LWTIGWFTKQNGKMSDILLQASVQVIDSTRCNADDAYQGEVTEKMCAGIPEGGVDTTC 230

310 LWTIGWFTKQNGKMSDILLQASVQVIDSTRCNADDAYQGEVTEKMCAGIPEGGVDTTC 369

231 QGDSGGPLMTQSDQHWVGVIVSGYGGGSGTSGVYTKVSAYLNWYVNWKAEL 284

370 QGDSGGPLMTQSDQHWVGVIVSGYGGGSGTSGVYTKVSAYLNWYVNWKAEL 423

RESULT 7
 ABG72428

ID ABG72428 standard; Protein; 423 AA.

XX AC ABG72428;

XX 05-FEB-2003 (first entry)

XX Human colorectal cancer modulating protein CUA8.

XX Colorectal cancer modulating protein; BCMP; CUA8; colorectal cancer; gene therapy; human.

XX Homo sapiens.

XX Key Location/Qualifiers
 FH Misc-difference 22
 FT /note= "Encoded by ATC"

XX US6455668-B1.

XX 24-SEP-2002.

XX 06-SEP-2000; 2000US-0656002.

XX 15-MAR-2000; 2000US-0525993.

XX 28-JAN-2000; 2000US-0493444.

XX (BOSB-) EOS BIOTECHNOLOGY INC.

XX Mack D, Gish KC, Wilson KE;

XX WPI; 2003-066245/06.

XX N-PSDB; ABS57763.

PT Diagnosing colorectal cancer comprises determining the expression of a gene encoding CUA8 in a first colon tissue of a first individual, and comparing them to the expression of a gene encoding CUA8 in a second normal tissue

PS Claim 15; Fig 2; 31pp; English.

CC The invention describes a method of diagnosing colorectal cancer comprising determining the expression of a gene encoding CUA8 or its fragment in a first colon tissue of an individual, and comparing the expression of the gene to the expression of the gene in normal tissue from the individual or in tissue from a second unaffected individual. A difference in the expression indicates that the first individual has colorectal cancer. The CUA8 genes and proteins are useful for diagnosing and prognosticating colorectal cancer. The methods are also useful for screening candidate bioactive agents that can modulate colorectal cancer, or for treating or inhibiting colorectal cancer in a patient e.g. by gene therapy. This is the amino acid sequence of the human colorectal cancer modulating protein CUA8.

XX Sequence 423 AA;

Query Match 79.1%; Score 1266; DB 24; Length 423;
 Best Local Similarity 97.4%; Pred. No. 7e-112; 4; Indels 0; Gaps 0;
 Matches 228; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 51 KIYGVYALVDVSNPQVSIQYDKQHVCGGSIIDPHWVLTAAHCFRKHDTVFNNKVRAGSD 110

190 RYVGGSEASVDSNPQVSIQYDKQHVCGGSIIDPHWVLTAAHCFRKHDTVFNNKVRAGSD 249

111 KLGSPSLAVAKIIIEFNPMYKNDIALMKLQPLTFSGTVRPICLPFFDELTPTATP 170

250 KLGSPSLAVAKIIIEFNPMYKNDIALMKLQPLTFSGTVRPICLPFFDELTPTATP 309

171 LWTIGWFTKQNGKMSDILLQASVQVIDSTRCNADDAYQGEVTEKMCAGIPEGGVDTTC 230

310 LWTIGWFTKQNGKMSDILLQASVQVIDSTRCNADDAYQGEVTEKMCAGIPEGGVDTTC 369

Y 231 QGDSGGFLMYQSDQWVGVGVSWGYGCGGSPSTPGVYTKVSAYLNWYVWKAEL 284
 b 370 QGDSGGFLMYQSDQWVGVGVSWGYGCGGSPSTPGVYTKVSAYLNWYVWKAEL 423

RESULT 8
 D ABU04918 standard; Protein; 423 AA.
 X C ABU04918;
 X T 29-JAN-2003 (first entry)
 X E Human expressed protein tag (EPT) #1584.
 X W Translational profiling; expressed protein tag; EPT; kinase;
 W phosphatase; protease; protease inhibitor; transporter;
 W cytoskeletal protein; receptor; transcription factor; cancer; MHC;
 W major histocompatibility complex; myeloma; colon cancer;
 W gastric cancer; adenocarcinoma; sarcoma; melanoma; lymphoma;
 W leukaemia.
 X S Homo sapiens.
 X N WO200278524-A2.
 X D 10-OCT-2002.
 X F 28-MAR-2002; 2002WO-US09671.
 X R 28-MAR-2001; 2001US-279495P.
 R 21-MAY-2001; 2001US-292544P.
 R 08-AUG-2001; 2001US-310801P.
 R 01-OCT-2001; 2001US-326370P.
 R 04-DEC-2001; 2001US-336780P.
 R 20-FEB-2002; 2002US-358985P.
 X A (ZYCO-) ZYCOS INC.
 X I Chicx RM, Tomlinson AJ, Urban RG;
 X WPI; 2003-040607/03.
 X R New polypeptides (e.g. kinases, phosphatases, proteases, transporters,
 T cytoskeletal proteins, receptors or transcription factors), useful for
 T treating cancer, e.g. colon cancer, gastric cancer, sarcoma, lymphoma
 T or leukemia -
 X S Example 2; SEQ ID No 1584; 134pp; English.
 X C The invention describes a purified polypeptide, which comprises a
 C fragment of a kinase, phosphatase, protease, protease inhibitor,
 C transporter, cytoskeletal protein, receptor or transcription factor.
 C The polypeptide is useful as an immunogenic composition for eliciting
 C in a mammal an immunogenic response directed against any of the purified
 C polypeptide. The purified polypeptide, or the antibody that binds to
 C this polypeptide, is useful for treating cancer. The polypeptide is
 C also useful for identifying compounds that binds to a naturally
 C processed class I or class II MHC-binding polypeptide. The polypeptides
 C and polynucleotides are particularly useful for treating or preventing
 C myeloma, colon cancer, gastric cancer, adenocarcinoma, sarcoma, melanoma,
 C lymphoma or leukaemia. These are also useful for screening agents for
 C treating the above mentioned diseases. This sequence represents an
 C expressed protein tag (EPT) isolated from human tissue for translational
 C profiling.
 C Note: This sequence does not appear in the printed specification but was
 C obtained in electronic format directly from WIPO at
 C ftp.wipo.int/pub/published_pct_sequences.
 X X Sequence 423 AA;

Query Match 79.1%; Score 1266; DB 24; Length 423;
 Best Local Similarity 97.4%; Pred. No. 7e-112;

Matches 228; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
 QY 51 KIVGGYALVDSDWPMQVSIQYDKQHVCGGSIILDPHWLTAHCFRKHTDVFNKVRAGSD 110
 Db 190 RVVGGEASVDNPMQVSIQYDKQHVCGGSIILDPHWLTAHCFRKHTDVFNKVRAGSD 249
 QY 111 KLGSPFLAVAKIIIIIEFNPMYPKNDIALMKLQPLTFSGTVRPICLFFPEELTPATP 170
 Db 250 KLGSPFLAVAKIIIIIEFNPMYPKNDIALMKLQPLTFSGTVRPICLFFPEELTPATP 309
 QY 171 LMTIHCFTKONGKMSDILLQASQVVIDSTRCNADDAVQGEVTERKMCAGIPEGGVDTTC 230
 Db 310 LMTIHCFTKONGKMSDILLQASQVVIDSTRCNADDAVQGEVTERKMCAGIPEGGVDTTC 369
 QY 231 QGDSGGFLMYQSDQWVGVGVSWGYGCGGSPSTPGVYTKVSAYLNWYVWKAEL 284
 Db 370 QGDSGGFLMYQSDQWVGVGVSWGYGCGGSPSTPGVYTKVSAYLNWYVWKAEL 423

RESULT 9
 ABU04919
 ID ABU04919 standard; Protein; 423 AA.
 XX AC ABU04919;
 XX DT 29-JAN-2003 (first entry)
 XX DE Human expressed protein tag (EPT) #1585.
 XX KW Translational profiling; expressed protein tag; EPT; kinase;
 KW phosphatase; protease; protease inhibitor; transporter;
 KW cytoskeletal protein; receptor; transcription factor; cancer; MHC;
 KW major histocompatibility complex; myeloma; colon cancer;
 KW gastric cancer; adenocarcinoma; sarcoma; melanoma; lymphoma;
 KW leukaemia.
 XX OS Homo sapiens.
 XX PN WO200278524-A2.
 XX PD 10-OCT-2002.
 XX PF 28-MAR-2002; 2002WO-US09671.
 XX PR 28-MAR-2001; 2001US-279495P.
 PR 21-MAY-2001; 2001US-292544P.
 PR 08-AUG-2001; 2001US-310801P.
 PR 01-OCT-2001; 2001US-326370P.
 PR 04-DEC-2001; 2001US-336780P.
 PR 20-FEB-2002; 2002US-358985P.
 XX (ZYCO-) ZYCOS INC.
 XX PA Chicx RM, Tomlinson AJ, Urban RG;
 XX WPI; 2003-040607/03.
 XX PT New polypeptides (e.g. kinases, phosphatases, proteases, transporters,
 PT cytoskeletal proteins, receptors or transcription factors), useful for
 PT treating cancer, e.g. colon cancer, gastric cancer, sarcoma, lymphoma
 PT or leukemia -
 XX PS Example 2; SEQ ID No 1585; 134pp; English.
 XX CC The invention describes a purified polypeptide, which comprises a
 CC fragment of a kinase, phosphatase, protease, protease inhibitor,
 CC transporter, cytoskeletal protein, receptor or transcription factor.
 CC The polypeptide is useful as an immunogenic composition for eliciting
 CC in a mammal an immunogenic response directed against any of the purified
 CC polypeptide. The purified polypeptide, or the antibody that binds to
 CC this polypeptide, is useful for treating cancer. The polypeptide is
 CC also useful for identifying compounds that binds to a naturally
 CC processed class I or class II MHC-binding polypeptide. The polypeptides

C and polynucleotides are particularly useful for treating or preventing
C myeloma, colon cancer, gastric cancer, adenocarcinoma, sarcoma, melanoma,
C lymphoma or leukaemia. These are also useful for screening agents for
C treating the above mentioned diseases. This sequence represents an
C expressed protein tag (EPT) isolated from human tissue for translational
C profiling.
C Note: This sequence does not appear in the printed specification but was
C obtained in electronic format directly from WIPO at
C ftp.wipo.int/pub/published_pct_sequences.

X Sequence 423 AA;

Query Match 79.1%; Score 1266; DB 24; Length 423;
Best Local Similarity 97.4%; Pred. No. 7e-112;
Matches 228; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
Y 51 KIYGGVALDVDSWFWQVSIQYDKQHVCGGSLDHPHWLTAHCFRKHDTVFNWVKRAGSD 110
b 190 RVYGGEEASVDSWFWQVSIQYDKQHVCGGSLDHPHWLTAHCFRKHDTVFNWVKRAGSD 249
Y 111 KLGSPFLAVAKIIIEFNPMYKNDIALMKLQPLTFSGTVRPICLPFFDEELTPATP 170
b 250 KLGSPFLAVAKIIIEFNPMYKNDIALMKLQPLTFSGTVRPICLPFFDEELTPATP 309
Y 171 LWIIGWFTKQNGKMSDILLOASVQVIDSTRCNADDAYQGEVTEKMKCAGIPGGVDTC 230
b 310 LWIIGWFTKQNGKMSDILLOASVQVIDSTRCNADDAYQGEVTEKMKCAGIPGGVDTC 369
Y 231 QGDSGGPLMYQSDQVHVGVISWVGCGGPGSTPGVYTKVSAYLNIYVWKAEL 284
b 370 QGDSGGPLMYQSDQVHVGVISWVGCGGPGSTPGVYTKVSAYLNIYVWKAEL 423

ESULT 10

BUD4923

D ABU04923 standard; Protein; 423 AA.

C ABU04923;

X 29-JAN-2003 (first entry)

T Human expressed protein tag (EPT) #1589.

X Translational profiling; expressed protein tag; EPT; kinase;

E phosphatase; protease; protease inhibitor; transporter;

X cytoskeletal protein; receptor; transcription factor; cancer; MHC;

W major histocompatibility complex; myeloma; colon cancer;

W gastric cancer; adenocarcinoma; sarcoma; melanoma; lymphoma;

W leukaemia.

W Homo sapiens.

S WO200278524-A2.

N 10-OCT-2002.

D 28-MAR-2002; 2002WO-US09671.

F 28-MAR-2001; 2001US-279495P.

X 21-MAY-2001; 2001US-292544P.

R 08-AUG-2001; 2001US-310801P.

R 01-OCT-2001; 2001US-326370P.

R 04-DEC-2001; 2001US-336780P.

R 20-FEB-2002; 2002US-358985P.

R (ZYCO-) ZYCOS INC.

A Chicx RM, Tomlinson AJ, Urban RG;

X WPI; 2003-040607/03.

X New polypeptides (e.g. kinases, phosphatases, proteases, transporters,

I cytoskeletal proteins, receptors or transcription factors), useful for

PT treating cancer, e.g. colon cancer, gastric cancer, sarcoma, lymphoma
PT or leukemia

XX Example 2; SEQ ID No 1589; 134pp; English.

XX The invention describes a purified polypeptide, which comprises a
CC fragment of a kinase, phosphatase, protease, protease inhibitor,
CC transporter, cytoskeletal protein, receptor or transcription factor.
CC The polypeptide is useful as an immunogenic composition for eliciting
CC in a mammal an immunogenic response directed against any of the purified
CC polypeptide. The purified polypeptide, or the antibody that binds to
CC this polypeptide, is useful for treating cancer. The polypeptide is
CC also useful for identifying compounds that binds to a naturally
CC processed class I or class II MHC-binding polypeptide. The polypeptides
CC and polynucleotides are particularly useful for treating or preventing
CC myeloma, colon cancer, gastric cancer, adenocarcinoma, sarcoma, melanoma,
CC lymphoma or leukaemia. These are also useful for screening agents for
CC treating the above mentioned diseases. This sequence represents an
CC expressed protein tag (EPT) isolated from human tissue for translational
CC profiling.

CC Note: This sequence does not appear in the printed specification but was
CC obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.

XQ Sequence 423 AA;

Query Match 79.1%; Score 1266; DB 24; Length 423;
Best Local Similarity 97.4%; Pred. No. 7e-112;
Matches 228; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 51 KIYGGVALDVDSWFWQVSIQYDKQHVCGGSLDHPHWLTAHCFRKHDTVFNWVKRAGSD 110
Db 190 RVYGGEEASVDSWFWQVSIQYDKQHVCGGSLDHPHWLTAHCFRKHDTVFNWVKRAGSD 249
QY 111 KLGSPFLAVAKIIIEFNPMYKNDIALMKLQPLTFSGTVRPICLPFFDEELTPATP 170
Db 250 KLGSPFLAVAKIIIEFNPMYKNDIALMKLQPLTFSGTVRPICLPFFDEELTPATP 309
QY 171 LWIIGWFTKQNGKMSDILLOASVQVIDSTRCNADDAYQGEVTEKMKCAGIPGGVDTC 230
Db 310 LWIIGWFTKQNGKMSDILLOASVQVIDSTRCNADDAYQGEVTEKMKCAGIPGGVDTC 369
QY 231 QGDSGGPLMYQSDQVHVGVISWVGCGGPGSTPGVYTKVSAYLNIYVWKAEL 284
Db 370 QGDSGGPLMYQSDQVHVGVISWVGCGGPGSTPGVYTKVSAYLNIYVWKAEL 423

RESULT 11

RAY99417

ID AAY99417 standard; Protein; 432 AA.

XX AAY99417;

XX 08-AUG-2000 (first entry)

XX Human PRO1570 (UNQ776) amino acid sequence SEQ ID NO:275.

DE Human; PRO polypeptide; membrane bound protein; receptor; diagnosis;

XX transmembrane; secretion; immunoadhesion; pharmaceutical; screening.

XX Homo sapiens.

XX WO200012708-A2.

XX 09-MAR-2000.

XX 01-SEP-1999; 99WO-US20111.

XX 01-SEP-1998; 98US-0098716.

XX 01-SEP-1998; 98US-0098749.

XX 01-SEP-1998; 98US-0098750.

XX 02-SEP-1998; 98US-0098803.

XX 02-SEP-1998; 98US-0098821.

R	02-SEP-1998;	98US-0038843.	PR	20-OCT-1998;	98US-0105002.
R	09-SEP-1998;	98US-0039536.	PR	21-OCT-1998;	98US-0105104.
R	09-SEP-1998;	98US-0039596.	PR	22-OCT-1998;	98US-0105169.
R	09-SEP-1998;	98US-0039598.	PR	22-OCT-1998;	98US-0105266.
R	09-SEP-1998;	98US-0039602.	PR	26-OCT-1998;	98US-0105693.
R	09-SEP-1998;	98US-0039642.	PR	26-OCT-1998;	98US-0105694.
R	10-SEP-1998;	98US-0099741.	PR	27-OCT-1998;	98US-0105807.
R	10-SEP-1998;	98US-0099754.	PR	27-OCT-1998;	98US-0105881.
R	10-SEP-1998;	98US-0099763.	PR	27-OCT-1998;	98US-0105882.
R	10-SEP-1998;	98US-0099792.	PR	27-OCT-1998;	98US-0106062.
R	10-SEP-1998;	98US-0099808.	PR	28-OCT-1998;	98US-0106023.
R	10-SEP-1998;	98US-0099812.	PR	28-OCT-1998;	98US-0106029.
R	10-SEP-1998;	98US-0099815.	PR	28-OCT-1998;	98US-0106030.
R	10-SEP-1998;	98US-0099816.	PR	28-OCT-1998;	98US-0106032.
R	15-SEP-1998;	98US-0100385.	PR	28-OCT-1998;	98US-0106033.
R	15-SEP-1998;	98US-0100388.	PR	28-OCT-1998;	98US-0106178.
R	15-SEP-1998;	98US-0100390.	PR	29-OCT-1998;	98US-0106248.
R	16-SEP-1998;	98US-0100584.	PR	29-OCT-1998;	98US-0106384.
R	16-SEP-1998;	98US-0100627.	PR	29-OCT-1998;	98US-0108500.
R	16-SEP-1998;	98US-0100661.	PR	30-OCT-1998;	98US-0108500.
R	16-SEP-1998;	98US-0100662.	PR	30-OCT-1998;	98US-0108500.
R	16-SEP-1998;	98US-0100664.	PR	03-NOV-1998;	98US-0106856.
R	17-SEP-1998;	98US-0100683.	PR	03-NOV-1998;	98US-0106902.
R	17-SEP-1998;	98US-0100684.	PR	03-NOV-1998;	98US-0106905.
R	17-SEP-1998;	98US-0100710.	PR	03-NOV-1998;	98US-0108919.
R	17-SEP-1998;	98US-0100711.	PR	03-NOV-1998;	98US-0108932.
R	17-SEP-1998;	98US-0100913.	PR	10-NOV-1998;	98US-0106934.
R	17-SEP-1998;	98US-0100930.	PR	10-NOV-1998;	98US-0107783.
R	18-SEP-1998;	98US-0100848.	PR	17-NOV-1998;	98US-0108775.
R	18-SEP-1998;	98US-0100849.	PR	17-NOV-1998;	98US-0108779.
R	18-SEP-1998;	98US-0101014.	PR	17-NOV-1998;	98US-0108787.
R	18-SEP-1998;	98US-0101068.	PR	17-NOV-1998;	98US-0108788.
R	18-SEP-1998;	98US-0101071.	PR	17-NOV-1998;	98US-0108801.
R	22-SEP-1998;	98US-0101279.	PR	17-NOV-1998;	98US-0108802.
R	23-SEP-1998;	98US-0101471.	PR	17-NOV-1998;	98US-0108806.
R	23-SEP-1998;	98US-0101472.	PR	17-NOV-1998;	98US-0108807.
R	23-SEP-1998;	98US-0101474.	PR	17-NOV-1998;	98US-0108867.
R	23-SEP-1998;	98US-0101475.	PR	17-NOV-1998;	98US-0108925.
R	23-SEP-1998;	98US-0101476.	PR	18-NOV-1998;	98US-0108848.
R	23-SEP-1998;	98US-0101477.	PR	18-NOV-1998;	98US-0108849.
R	23-SEP-1998;	98US-0101479.	PR	18-NOV-1998;	98US-0108850.
R	24-SEP-1998;	98US-0101738.	PR	18-NOV-1998;	98US-0108851.
R	24-SEP-1998;	98US-0101741.	PR	18-NOV-1998;	98US-0108852.
R	24-SEP-1998;	98US-0101743.	PR	18-NOV-1998;	98US-0108858.
R	24-SEP-1998;	98US-0101915.	PR	18-NOV-1998;	98US-0108858.
R	29-SEP-1998;	98US-0102207.	PR	18-NOV-1998;	98US-0108904.
R	29-SEP-1998;	98US-0102240.	PR	18-NOV-1998;	98US-0108904.
R	29-SEP-1998;	98US-0102307.	PR	18-NOV-1998;	98US-0108904.
R	29-SEP-1998;	98US-0102330.	PR	18-NOV-1998;	98US-0108904.
R	29-SEP-1998;	98US-0102331.	PR	18-NOV-1998;	98US-0108904.
R	30-SEP-1998;	98US-0102484.	PR	18-NOV-1998;	98US-0108904.
R	30-SEP-1998;	98US-0102487.	PR	18-NOV-1998;	98US-0108904.
R	30-SEP-1998;	98US-0102570.	PR	18-NOV-199	

```
Y 51 KIVGGYALDVDSWPNQVSIQYDKQHVCGSSILDPHWVLTAAHCFRKHDTVDVFNKVRAGSD 110
b :|||
199 RVVGGEASVDSWPNQVSIQYDKQHVCGSSILDPHWVLTAAHCFRKHDTVDVFNKVRAGSD 258
Y 111 KLGSPFLAVAKIIIEFNPNPKNDIALMKLOPPLTFSGTVRPICLPFFDEELTPATP 170
b :|||
259 KLGSPFLAVAKIIIEFNPNPKNDIALMKLOPPLTFSGTVRPICLPFFDEELTPATP 318
Y 171 LWIIGWFTKQNGKMSDILLQASVQVIDSTRCNADDAYQGEVTEKMCACIPEGGVDTTC 230
b :|||
319 LWIIGWFTKQNGKMSDILLQASVQVIDSTRCNADDAYQGEVTEKMCACIPEGGVDTTC 378
Y 231 QGDSGGPLMYQSDQHWVGVIVSWGVCGGPSTPGVYTKVSAYLNIYNVWKAE 284
b :|||
379 QGDSGGPLMYQSDQHWVGVIVSWGVCGGPSTPGVYTKVSAYLNIYNVWKAE 432

RESULT 12
AU29188
D AAU29188 standard; Protein; 432 AA.
X C
X C AAU29188;
X T
X T 18-DEC-2001 (first entry)
X E Human PRO polypeptide sequence #165.
X X
W PRO polypeptide; mammal; tumour; cancer; human; cattle; horse; sheep;
W dog; cat; pig; goat; rabbit; tumour necrosis factor alpha; TNF-alpha;
W blood; chondrocyte cell; cell proliferation; cell differentiation; colon;
W adrenal; lung; breast; prostate; rectum; cervix; liver; genetic disorder.
X X
S Homo sapiens.
X X
N WO200168848-A2.
X D
X D 20-SEP-2001.
X F
X F 28-FEB-2001; 2001WO-US06520.
X X
R 01-MAR-2000; 2000WO-US05601.
R 02-MAR-2000; 2000WO-US05841.
R 03-MAR-2000; 2000US-187202P.
R 06-MAR-2000; 2000US-186968P.
R 14-MAR-2000; 2000US-189320P.
R 14-MAR-2000; 2000US-189328P.
R 15-MAR-2000; 2000WO-US06884.
R 21-MAR-2000; 2000US-190828P.
R 21-MAR-2000; 2000US-191007P.
R 21-MAR-2000; 2000US-191048P.
R 21-MAR-2000; 2000US-191314P.
R 28-MAR-2000; 2000US-192655P.
R 29-MAR-2000; 2000US-193032P.
R 30-MAR-2000; 2000US-193053P.
R 30-MAR-2000; 2000WO-US08439.
R 04-APR-2000; 2000US-194439P.
R 04-APR-2000; 2000US-194647P.
R 11-APR-2000; 2000US-195975P.
R 11-APR-2000; 2000US-196000P.
R 11-APR-2000; 2000US-196187P.
R 11-APR-2000; 2000US-196690P.
R 11-APR-2000; 2000US-196820P.
R 18-APR-2000; 2000US-198121P.
R 18-APR-2000; 2000US-198585P.
R 25-APR-2000; 2000US-199397P.
R 25-APR-2000; 2000US-199550P.
R 25-APR-2000; 2000US-199654P.
R 03-MAY-2000; 2000US-201516P.
R 17-MAY-2000; 2000WO-US13705.
R 22-MAY-2000; 2000WO-US14042.
R 30-MAY-2000; 2000WO-US14941.
R 02-JUN-2000; 2000WO-US15264.
```

```
PR 05-JUN-2000; 2000US-209832P.
PR 28-JUL-2000; 2000WO-US20710.
PR 22-AUG-2000; 2000US-0644848.
PR 24-AUG-2000; 2000WO-US23328.
PR 08-NOV-2000; 2000WO-US30952.
PR 01-DEC-2000; 2000WO-US32678.
PR 20-DEC-2000; 2000WO-US34956.
XX (GETH ) GENENTECH INC.
XX Baker KP, Chen J, Desnoyers L, Goddard A, Godowski PJ, Gurney AL;
XX Pan J, Smith V, Watanabe CK, Wood WI, Zhang Z;
XX WPI; 2001-602746/68.
XX DR N-PSDE; AAS46089.
XX Novel nucleic acids encoding PRO polypeptides, used to diagnose the
XX presence of tumours, such as prostate and breast tumours, in mammals and
XX to screen for modulators of the compounds -
XX Claim 11; Fig 330; 774pp; English.
XX Sequences AAU29024-AAU29328 represent PRO polypeptides of the invention.
XX The PRO polypeptides and their associated nucleic acids can be used to
XX detect the presence of a tumour in a mammal by comparing the level of
XX expression of a PRO polypeptide in a test sample of cells from the animal
XX and a control sample of normal cells, whereby a higher level of
XX expression in the test sample indicates the presence of a tumour in the
XX mammal. Mammals include dogs, cats, cattle, horses, sheep, pigs, goats
XX and rabbits but are preferably human. The polypeptides can be used to
XX stimulate tumour necrosis factor (TNF) alpha release from human blood,
XX when contacted with it. A specific polypeptide can be used to stimulate
XX the proliferation or differentiation of chondrocyte cells. The PRO
XX proteins can be used to determine the presence of tumours and also
XX susceptibility to tumour development, particularly adrenal, lung, colon,
XX breast, prostate, rectal, cervical, or liver tumours, in mammalian
XX subjects. The oligonucleotide probes specific for the PRO nucleic acids
XX can be used for genetic analysis of individuals with genetic disorders.
XX SQ Sequence 432 AA;
```

```
Query Match 79.1%; Score 1266; DB 22; Length 432;
Best Local Similarity 97.4%; Pred. No. 7.1e-112;
Matches 228; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
QY 51 KIVGGYALDVDSWPNQVSIQYDKQHVCGSSILDPHWVLTAAHCFRKHDTVDVFNKVRAGSD 110
Db :|||
199 RVVGGEASVDSWPNQVSIQYDKQHVCGSSILDPHWVLTAAHCFRKHDTVDVFNKVRAGSD 258
QY 111 KLGSPFLAVAKIIIEFNPNPKNDIALMKLOPPLTFSGTVRPICLPFFDEELTPATP 170
Db :|||
259 KLGSPFLAVAKIIIEFNPNPKNDIALMKLOPPLTFSGTVRPICLPFFDEELTPATP 318
QY 171 LWIIGWFTKQNGKMSDILLQASVQVIDSTRCNADDAYQGEVTEKMCACIPEGGVDTTC 230
Db :|||
319 LWIIGWFTKQNGKMSDILLQASVQVIDSTRCNADDAYQGEVTEKMCACIPEGGVDTTC 378
QY 231 QGDSGGPLMYQSDQHWVGVIVSWGVCGGPSTPGVYTKVSAYLNIYNVWKAE 284
Db :|||
379 QGDSGGPLMYQSDQHWVGVIVSWGVCGGPSTPGVYTKVSAYLNIYNVWKAE 432
```

```
RESULT 13
AAB87581
ID AAB87581 standard; Protein; 432 AA.
XX
AC AAB87581;
XX
DT 15-MAY-2001 (first entry)
XX
DE Human PRO1570.
XX
KW Human; PRO protein; mapping.
```

X S Homo sapiens.
X N WO200116318-A2.
X D 08-MAR-2001.
X F 24-AUG-2000; 2000WO-US23328.
X R 01-SEP-1999; 99WO-US20111.
X R 15-SEP-1999; 99WO-US21090.
X R 07-DEC-1999; 99US-0159495.
X R 09-DEC-1999; 99US-0170262.
X R 11-JAN-2000; 2000US-0175481.
X R 18-FEB-2000; 2000WO-US04341.
X R 18-FEB-2000; 2000WO-US04342.
X R 22-FEB-2000; 2000WO-US04414.
X R 01-MAR-2000; 2000WO-US05601.
X R 03-MAR-2000; 2000US-017202.
X R 25-APR-2000; 2000US-0199397.
X R 22-MAY-2000; 2000WO-US14042.
X R 05-JUN-2000; 2000US-0209832.
X A (GETH) GENENTECH INC.
X I Eaton DL, Filvaroff E, Gerritsen ME, Goddard A, Godowski PU;
X I Grimaldi CJ, Gurney AL, Watanabe CK, Wood WI;
X X WPI; 2001-183260/18.
X R N-PSDB; AAF92113.
X T Eighty four nucleic acids encoding PRO polypeptides, useful in
T molecular biology, including use as hybridization probes, and in
T chromosome and gene mapping.
X X Claim 12; Fig 112; 278pp; English.
X S The present sequence is a human PRO polypeptide (secreted and
C transmembrane). The PRO protein, and PRO agonists, PRO antagonists or
C anti-PRO antibodies are useful for preparation of a medicament useful in
C the treatment of a condition which is responsive to the PRO protein,
C agonists, antagonists or anti-PRO antibodies. The PRO protein may also be
C employed as molecular weight markers for protein electrophoresis. The PRO
C coding sequence has applications in molecular biology, including use as
C hybridisation probes, and in chromosome and gene mapping.
X Q Sequence 432 AA;
Query Match 79.1%; Score 1266; DB 22; Length 432;
Best Local Similarity 97.4%; Pred. No. 7.1e-112;
Matches 228; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
Y 51 KIVGGYALVDSPWQVSIQYDKQHVCGGSIIDPHWVLTAAHCFRKHDTVFNKVRAGSD 110
b 199 RVVGGEEASVDSWQVSIQYDKQHVCGGSIIDPHWVLTAAHCFRKHDTVFNKVRAGSD 258
Y 111 KLGSPFSLAVAKIIIEFNPMYPKNDIALMKLQPLTFSGTVRPICLPFFDEELTPATP 170
b 259 KLGSPFSLAVAKIIIEFNPMYPKNDIALMKLQPLTFSGTVRPICLPFFDEELTPATP 318
Y 171 LMIIGWFTKQNGKMSDILLOASVQVIDSTRCNADDAAYQGEVTEKMCAGIPEGGVDT 230
b 319 LMIIGWFTKQNGKMSDILLOASVQVIDSTRCNADDAAYQGEVTEKMCAGIPEGGVDT 378
Y 231 QGDSGGFLMYQSDQHWVGVISWGYCGGSPSTPGVYTKVSAYLNWIYNVWKAEL 284
b 379 QGDSGGFLMYQSDQHWVGVISWGYCGGSPSTPGVYTKVSAYLNWIYNVWKAEL 432
RESULT 14
AB66166
D AAB66166 standard; protein; 432 AA.
X

AC AAB66166;
XX 02-APR-2001 (first entry)
DT Protein of the invention #78.
DE Secreted; transmembrane; gene therapy.
XX Unidentified.
XX WO200078961-A1.
XX 28-DEC-2000.
XX 18-FEB-2000; 2000WO-US04342.
XX 23-JUN-1999; 99US-0141037.
XX 20-JUL-1999; 99US-0144758.
XX 26-JUL-1999; 99US-0145698.
XX 01-SEP-1999; 99WO-US20111.
XX 29-OCT-1999; 99US-0162506.
XX 30-NOV-1999; 99WO-US28313.
XX 02-DEC-1999; 99WO-US28551.
XX 16-DEC-1999; 99WO-US30095.
XX 05-JAN-2000; 2000WO-US00219.
XX 06-JAN-2000; 2000WO-US00376.
XX (GETH) GENENTECH INC.
XX Baker KP, Botstein D, Desnoyers L, Eaton DL, Ferrara N, Fong S;
XX Gao W, Goddard A, Godowski PJ, Grimaldi CJ, Gurney AL, Hillan KJ;
XX Pan J, Paoni NF, Roy MA, Smith V, Stewart TA, Tumas D;
XX Watanabe CK, Williams PM, Wood WI;
XX WPI; 2001-071395/08.
XX Secreted and transmembrane proteins and nucleic acids designated PRO,
PT useful as hybridization probes, in chromosome and gene mapping and gene
PT therapy -
XX Claim 1; Fig 156; 787pp; English.
XX The present invention relates to secreted and transmembrane proteins.
CC These proteins and the DNA encoding them may be used as hybridization
CC probes, in chromosome and gene mapping and in the generation of
CC anti-sense RNA and DNA. They may also be used to generate either
CC transgenic animals or knockout animals which are in turn useful for
CC development and screening of therapeutically useful reagents.
CC The nucleic acids may also be used in gene therapy.
XX SQ Sequence 432 AA;
Query Match 79.1%; Score 1266; DB 22; Length 432;
Best Local Similarity 97.4%; Pred. No. 7.1e-112;
Matches 228; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
QY 51 KIVGGYALVDSPWQVSIQYDKQHVCGGSIIDPHWVLTAAHCFRKHDTVFNKVRAGSD 110
Db 199 RVVGGEEASVDSWQVSIQYDKQHVCGGSIIDPHWVLTAAHCFRKHDTVFNKVRAGSD 258
QY 111 KLGSPFSLAVAKIIIEFNPMYPKNDIALMKLQPLTFSGTVRPICLPFFDEELTPATP 170
Db 259 KLGSPFSLAVAKIIIEFNPMYPKNDIALMKLQPLTFSGTVRPICLPFFDEELTPATP 318
QY 171 LMIIGWFTKQNGKMSDILLOASVQVIDSTRCNADDAAYQGEVTEKMCAGIPEGGVDT 230
Db 319 LMIIGWFTKQNGKMSDILLOASVQVIDSTRCNADDAAYQGEVTEKMCAGIPEGGVDT 378
QY 231 QGDSGGFLMYQSDQHWVGVISWGYCGGSPSTPGVYTKVSAYLNWIYNVWKAEL 284
Db 379 QGDSGGFLMYQSDQHWVGVISWGYCGGSPSTPGVYTKVSAYLNWIYNVWKAEL 432

RESULT 15

ABG95906
 ABG95906 standard; Protein; 432 AA.

ABG95906;

10-DEC-2002 (first entry)

Human secreted/transmembrane protein PRO1570.

Human; secreted protein; transmembrane protein; antirheumatic;
 antiarthritic; osteopathic; sports-related joint problem;
 articular cartilage defect; osteoarthritis; rheumatoid arthritis.

Homo sapiens.

US2002119130-A1.

29-AUG-2002.

06-DEC-2001; 2001US-0006867.

29-OCT-1997; 97US-063435P.

29-OCT-1997; 97US-064215P.

22-APR-1998; 98US-082797P.

29-APR-1998; 98US-083495P.

15-MAY-1998; 98US-085579P.

10-JUN-1998; 98US-088811P.

10-JUN-1998; 98US-088824P.

10-JUN-1998; 98US-088825P.

11-JUN-1998; 98US-088863P.

12-JUN-1998; 98US-089105P.

16-JUN-1998; 98US-089514P.

16-SEP-1998; 98WO-US19330.

08-MAR-1999; 99WO-US10733.

14-MAY-1999; 99WO-US12733.

02-JUN-1999; 99WO-US12252.

01-SEP-1999; 99WO-US20111.

15-SEP-1999; 99WO-US21090.

15-SEP-1999; 99WO-US21194.

22-DEC-1999; 99WO-US30720.

18-FEB-2000; 2000WO-US04341.

18-FEB-2000; 2000WO-US04342.

30-MAR-2000; 2000WO-US08439.

22-MAY-2000; 2000WO-US14042.

02-JUN-2000; 2000WO-US15264.

23-AUG-2000; 2000WO-US23522.

24-AUG-2000; 2000WO-US23328.

10-NOV-2000; 2000WO-US30873.

01-DEC-2000; 2000WO-US32378.

20-DEC-2000; 2000WO-US34956.

28-FEB-2001; 2001WO-US06520.

20-JUN-2001; 2001WO-US19692.

29-JUN-2001; 2001WO-US21066.

09-JUL-2001; 2001WO-US21735.

(GETH) GENENTECH INC.

WPI; 2002-731348/79.

N-PSDB; ABS74433.

New isolated secreted and transmembrane PRO polypeptide useful for
 modulating biological activity of a cell, or for treating
 sports-related joint problems, osteoarthritis or rheumatoid arthritis

Claim 20; Fig 112; 399pp; English.

The invention relates to an isolated secreted and transmembrane PRO
 polypeptide having 80 % sequence identity to a sequence appearing
 as ABG95951-ABG95934 or their associated signal peptide, or a sequence of

an extracellular domain of the proteins with their associated signal
 peptide or lacking its associated signal peptide. Also included are
 the nucleic acids encoding the proteins, vectors, host cells,
 fusion proteins and antibodies which specifically bind to the proteins.
 The proteins are useful for detecting a polypeptide designated as A, B, C
 or D in a sample suspected of containing an A, B, C or D polypeptide,
 by contacting the sample with a polypeptide designated as E, F, G, H or
 I (or vice versa) and determining the formation of a A/E, B/F, C/H
 or D/I polypeptide conjugate in the sample, where the formation of the
 conjugate is indicative of the presence of an A, B, C or D polypeptide
 in the sample, where A is a PRO10272 polypeptide, B is a PRO20110
 polypeptide, C is a PRO10096 polypeptide, D is a PRO19760 polypeptide,
 E is a PRO5801 polypeptide, F is a PRO1 polypeptide, G is a PRO20040
 polypeptide, H is a PRO20233 polypeptide and I is a PRO1890
 polypeptide. The sample comprises a cell suspected of expressing the A,
 B, C or D polypeptide. The E, F, G, H or I polypeptide is labeled with
 a detectable label or is attached to a solid support. The proteins are
 useful for linking a bioactive molecule to a cell expressing a
 polypeptide designated as A, B, C or D or E, F, G, H or I. The bioactive
 molecule is a toxin, a radiolabel or an antibody. The bioactive molecule
 causes death of the cell. A, B, C, D, E, F, G, H, or I, or antibodies
 against them are useful for modulating a biological activity of a cell
 expressing a polypeptide designated as A, B, C or D or E, F, G, H, or
 I. The cell is killed. The proteins are useful for identifying
 agonists or antagonists, for the preparation of a medicament useful in
 the treatment of a condition which is responsive to the proteins, as
 molecular weight markers for protein electrophoresis purposes, and as
 therapeutic agents for treating sports-related joint problems,
 articular cartilage defects, osteoarthritis or rheumatoid arthritis.
 Nucleic acids encoding the proteins are useful as hybridisation probes,
 in chromosome and gene mapping, in the generation of anti-sense RNA and
 DNA, for the preparation of the proteins, to generate transgenic or
 knockout animals which are useful in the development and screening of
 therapeutic useful reagents, for chromosome identification, and in gene
 therapy. The antibody is useful as a therapeutic agent, in a diagnostic
 assay and for affinity purification of the protein from recombinant
 cell culture natural sources. The present sequence represents a novel
 secreted or transmembrane protein of the invention.

XX Sequence 432 AA;

Query Match 79.1%; Score 1266; DB 23; Length 432;

Best Local Similarity 97.4%; Pred. NO. 7.1e-112;

Matches 228; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 51 KIVGVALDVDSWPMQVSIQYDKQHVCGGSIIDPHWVLTAAHCFRKHDTDFNNKVRAGSD 110

Db 199 RVVGGEASVDSWPMQVSIQYDKQHVCGGSIIDPHWVLTAAHCFRKHDTDFNNKVRAGSD 258

QY 111 KLGSPSLAVAKIIITIEFNPMYPKNDIALMKLQFPLTSGTVRPICLPFFDEELTPATP 170

Db 259 KLGSPSLAVAKIIITIEFNPMYPKNDIALMKLQFPLTSGTVRPICLPFFDEELTPATP 318

QY 171 LWIIGWGFTKQNGKMSDILLOASVOVIDSTRCNADDAVCGEVTCKMCGAGIPEGGVDTG 230

Db 319 LWIIGWGFTKQNGKMSDILLOASVOVIDSTRCNADDAVCGEVTCKMCGAGIPEGGVDTG 378

QY 231 QGDSGGPLMYQSDQHWVGIVSWGCGGPGSPGVYTKVSAYLNWLYNNWKAL 284

Db 379 QGDSGGPLMYQSDQHWVGIVSWGCGGPGSPGVYTKVSAYLNWLYNNWKAL 432

Search completed: November 23, 2003, 07:49:06

Job time : 43.9766 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

M protein - protein search, using sw model

un on: November 23, 2003, 07:47:10 ; Search time 14.8611 Seconds
(without alignments)
831.351 Million cell updates/sec

title: US-09-607-745-9

effect score: 1600

sequence: 1 MDSKSSQKSLLLLLLVSN.....LNWLNWVKALSRHHHHH 292

coring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

searched: 328717 seqs, 42310858 residues

total number of hits satisfying chosen parameters: 328717

inimum DB seq length: 0

aximum DB seq length: 2000000000

ost-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

atabase : Issued Patents AA.*

- 1: /cgn2_6/ptodata/2/iaa/5A.COMB.pcp.*
- 2: /cgn2_6/ptodata/2/iaa/5B.COMB.pcp.*
- 3: /cgn2_6/ptodata/2/iaa/6A.COMB.pcp.*
- 4: /cgn2_6/ptodata/2/iaa/6B.COMB.pcp.*
- 5: /cgn2_6/ptodata/2/iaa/PTUS.COMB.pcp.*
- 6: /cgn2_6/ptodata/2/iaa/backfiles.pcp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

result No.	Score	Query Match	Length	DB ID	Description
1	1266	79.1	423	4	US-09-656-002-2
2	1266	79.1	435	3	US-09-008-271A-6
3	765	47.8	316	4	US-09-387-375-9
4	754.5	47.2	315	4	US-09-386-653A-9
5	746	46.6	328	4	US-09-386-642-11
6	740.5	46.3	327	4	US-09-386-629-8
7	672.5	42.0	289	4	US-09-386-642-14
8	645	40.3	288	4	US-09-386-642-13
9	601.5	37.6	306	4	US-09-386-642-53
10	593	37.1	319	4	US-09-386-642-12
11	574	35.9	284	4	US-09-386-642-54
12	526.5	32.9	418	1	US-08-508-448C-25
13	526.5	32.9	418	4	US-09-370-838-82
14	526.5	32.9	418	4	US-09-370-838-83
15	524.5	32.8	418	4	US-09-370-838-62
16	522.5	32.7	232	1	US-08-508-448C-19
17	507	31.7	454	3	US-09-518-046-2
18	506	31.6	492	3	US-09-342-749-2
19	506	31.6	492	4	US-09-691-840-2
20	505.5	31.6	492	4	US-09-820-002-2
21	505.5	31.6	417	4	US-09-820-002-4
22	499.5	31.2	256	2	US-09-027-337-3
23	499.5	31.2	256	4	US-09-644-600-3
24	497.5	31.1	255	3	US-08-944-483-67
25	497	31.1	283	3	US-08-807-151-1
26	497	31.1	283	4	US-09-478-957-1
27	491.5	30.7	416	2	US-09-000-846-2

28 489 30.6 638 2 US-08-681-151-3 Sequence 3, Appli
29- 482 30.1 285 4 US-09-023-942A-26 Sequence 26, Appli
30 477 29.8 248 3 US-08-944-483-63 Sequence 63, Appli
31 463.5 29.0 246 2 US-08-978-404B-44 Sequence 44, Appli
32 463.5 29.0 250 3 US-08-944-483-68 Sequence 68, Appli
33 461.5 28.8 407 4 US-09-734-675-4 Sequence 4, Appli
34 456.5 28.5 314 4 US-09-023-942A-6 Sequence 6, Appli
35 455 28.4 284 4 US-09-387-375-7 Sequence 7, Appli
36 453.5 28.3 247 2 US-08-956-267A-2 Sequence 2, Appli
37 451.5 28.2 314 3 US-09-008-271A-3 Sequence 3, Appli
38 450.5 28.2 312 4 US-09-023-942A-4 Sequence 4, Appli
39 444 27.8 855 2 US-09-027-337-2 Sequence 2, Appli
40 444 27.8 855 4 US-08-644-600-2 Sequence 2, Appli
41 443 27.7 790 4 US-08-991-761A-13 Sequence 13, Appli
42 442.5 27.7 290 4 US-09-386-653A-7 Sequence 7, Appli
43 441.5 27.6 812 1 US-08-248-629A-1 Sequence 1, Appli
44 441.5 27.6 812 1 US-08-451-932-1 Sequence 1, Appli
45 441.5 27.6 812 1 US-08-452-260-1 Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-09-656-002-2
; Sequence 2, Application US/09656002
; Patent No. 6455668
; GENERAL INFORMATION:
; APPLICANT: Mack, David
; APPLICANT: Gish, Kurt
; APPLICANT: Wilson, Keith
; TITLE OF INVENTION: NOVEL METHODS OF DIAGNOSING COLORECTAL CANCER, COMPOSITIONS, AND N
; TITLE OF INVENTION: OF SCREENING FOR COLORECTAL CANCER MODULATORS
; FILE REFERENCE: A-69108/DJB/JJD/AMS
; CURRENT APPLICATION NUMBER: US/09/656,002
; CURRENT FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: US 09/525,993
; PRIOR FILING DATE: 2000-03-15
; PRIOR APPLICATION NUMBER: US 09/493,444
; PRIOR FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: PCT/US 00/07044
; PRIOR FILING DATE: 2000-03-15
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 423
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-656-002-2

Query Match 79.1%; Score 1266; DB 4; Length 423;
Best Local Similarity 97.4%; Pred. No. 5.5e-128;
Matches 228; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy 51 KIYGVYALVDSPWQVSIQYDKQHVCGSILDPHWLTAHCFKTKDYFNWVKRAGSD 110
Db 190 RVVGEASVDSWPQVSIQYDKQHVCGSILDPHWLTAHCFKTKDYFNWVKRAGSD 249
Qy 111 KLGSPFLAVAKIIIEFNPMYKNDIALMKLQFPLTFSGTVRPICLPRFDEELTPATP 170
Db 250 KLGSPFLAVAKIIIEFNPMYKNDIALMKLQFPLTFSGTVRPICLPRFDEELTPATP 309
Qy 171 LWIIGWFTKQNGKMSDILLQASVQVIDSTRCNADDAYQGVTEKMKCAGIEPEGGYDTC 230
Db 310 LWIIGWFTKQNGKMSDILLQASVQVIDSTRCNADDAYQGVTEKMKCAGIEPEGGYDTC 369
Qy 231 QGSGGFLMTQSDQHWVGVISWGYCGGSPSTPGVYTKVSAYLNWYNWKAEL 284
Db 370 QGSGGFLMTQSDQHWVGVISWGYCGGSPSTPGVYTKVSAYLNWYNWKAEL 423

RESULT 2
US-09-008-271A-6

Query Match 47.2%; Score 754.5; DB 4; Length 315;
Best Local Similarity 48.4%; Pred. No. 4.6e-73;
Matches 156; Conservative 46; Mismatches 83; Indels 37; Gaps 10;

Y 1 MDSKGSQKSRLLLLVSVNLLCGVVDYKDDDDVDAALAAPFDDDDKIVGGYALDV 60
b 1 MDSKGSQKSRLLLLVSVNLLCGVVDYKDDDDVDAALAAPFDDDDKIVGGYALEE 60

Y 61 DSWPQVSIQYDKQHVCGGSIIDPHWVLTAAHCFRKHDTDFVFNKVRAGSDKL-GSFPSLA 119
b 61 GEWPQVSIQVNGSHFCGSLIAEQWVLTAAHCFRKHDTDFVFNKVRAGSDKL-GSFPSLA 120

Y 120 VAKIIIEFNPMY---PKNDIALMKLOPFTFSGTVRPICLPFFDELTATPLI--WII 174
b 121 YARVQVESNPLYQGTASSADVALVEAPVPFTNYILFVCLP--DPSVIFETGNCVWT 178

Y 175 GWGFTKQNGKMSD-----ILLQASVOVIDSTRCN-----ADDAQGE-VTEKMCAG 221
b 179 GWGSPSE-----EDLLPRLQKLAFLIITPKCNLLYSKDTFEGYQPKTKNDMLCAG 233

Y 222 IPEGVDTCQDGSGLPLMYQSDQ-WHVVGIVSWGCGGSPSTPGVYTKVSAYLNIWV- 279
b 234 FEEGKACKGDSGLPLVCLVQSWLQAGVLSWGGCGARQNPQVYIVTAHNNWIRII 293

Y 280 -----WKAEISRRHHHHH 292
b 294 PKLOFQPARLGGQKSRHHHHH 315

RESULT 5
S-09-386-642-11
Sequence 11, Application US/09386642
Patent No. 6420157
GENERAL INFORMATION:
APPLICANT: Darrow, Andrew
APPLICANT: Qi, Jensen
APPLICANT: Andrade-Gordon, Patricia
TITLE OF INVENTION: Zymogen Activation System
FILE REFERENCE: ORT-1028
CURRENT APPLICATION NUMBER: US/09/386,642
CURRENT FILING DATE: 1999-08-31
NUMBER OF SEQ ID NOS: 60
SOFTWARE: Patent in Ver. 2.0
SEQ ID NO 11
LENGTH: 328
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Fusion gene
OTHER INFORMATION: with homo sapien serine protease catalytic domain
US-09-386-642-11

Query Match 46.6%; Score 746; DB 4; Length 328;
Best Local Similarity 49.7%; Pred. No. 4e-72; Indels 36; Gaps 8;
Matches 163; Conservative 35; Mismatches 94; Indels 36; Gaps 8;

Y 1 MDSKGSQKSRLLLLVSVNLLCGVVDYKDDDDVDAALAAPFDDDDKIVGGYALDV 60
b 1 MDSKGSQKSRLLLLVSVNLLCGVVDYKDDDDVDAALAAPFDDDDKIVGGYALEA 60

Y 61 DSWPQVSIQYDKQHVCGGSIIDPHWVLTAAHCFRKHDTDFVFNKVRAGSDKLGSFPSLA- 119
b 61 GQWPQVSIQVNGSHFCGSLIAEQWVLTAAHCFRKHDTDFVFNKVRAGSDKLGSFPSLA- 120

Y 120 VAKIIIEFNPMYKD---NDIALMKLOPFTFSGTVRPICLPFFDELTATPLIWCW 176
b 121 VSTLKDIIIPHSYLGQSGQGLIALLQSRITFSRYIRPICLPANASFPNGLHCTVGV 180

Y 177 GFTKQNGKMS--DILLQASVOVIDSTRCNA---DDAQGE---VTEKMCAGIPEGGVD 229
b 181 GHVAPSVLLTPKPLQLEVLISRETNCNLINIDAKPEEFHFOEDVDCAGYVEGGKA 240

QY 230 CQDSSGGLMYQSD-QWHVVGIVSWGCGGSPSTPGVYTKVSAYLNIWY----- 277
b 241 CQDSSGGLPLSCPEGLWLTGIVSWGCGGSPSTPGVYTKVSAYLNIWY----- 300

QY 278 -----NVWKAEL---SRHHHHH 292
b 301 PQTESQPDNLCSHLAFSSRRHHHHH 328

RESULT 6
US-09-386-629-8
Sequence 8, Application US/09386629
Patent No. 6426199
GENERAL INFORMATION:
APPLICANT: Darrow, Andrew L.
APPLICANT: Qi, Jensen
APPLICANT: Andrade-Gordon, Patricia
TITLE OF INVENTION: Identification and Characterization of the complementary
TITLE OF INVENTION: DNA encoding the novel human serine protease C-E
FILE REFERENCE: ORT-1030
CURRENT APPLICATION NUMBER: US/09/386,629
CURRENT FILING DATE: 1999-08-31
NUMBER OF SEQ ID NOS: 11
SOFTWARE: Patent in Ver. 2.0
SEQ ID NO 8
LENGTH: 327
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: C-E catalytic
OTHER INFORMATION: domain fusion protein
US-09-386-629-8

Query Match 46.3%; Score 740.5; DB 4; Length 327;
Best Local Similarity 49.8%; Pred. No. 1.6e-71;
Matches 164; Conservative 39; Mismatches 87; Indels 39; Gaps 10;

QY 1 MDSKGSQKSRLLLLVSVNLLCGVVDYKDDDDVDAALAAPFDDDDKIVGGYALDV 60
b 1 MDSKGSQKSRLLLLVSVNLLCGVVDYKDDDDVDAALAAPFDDDDKIVGGYALEE 60

QY 61 DSWPQVSIQYDKQHVCGGSIIDPHWVLTAAHCFRKHDTDFVFNKVRAGSDKLGSFPSLA 119
b 61 SEWPVSIQVNGSHFCGSLIAEQWVLTAAHCFRKHDTDFVFNKVRAGSDKLGSFPSLA 119

QY 120 VAKIIIEFNPMY-KDN---DIALMKLOPFTFSGTVRPICLPFFDELTATPLIWI 173
b 120 SQKGVAVWVPHVYVSWKGCADIALVRLERSICFSEVLPICLPDASIHLPNTHCWI 179

QY 174 IGMFTKQNGKM--SDILLQASVOVIDSTRCNA---DDAQGEVTEKMCAGIPEGVD 228
b 180 SGWG-STQDGVPLPFPQTLQKLVPIIDSEVCSHLYWRGAGQSPITEDMLCAGYLEGERD 238

QY 229 TCQDSSGGLMYQSD-QWHVVGIVSWGCGGSPSTPGVYTKVSAYLNIWVW- 281
b 239 ACJGDSGGLMVCQVDGAWLGIISWGGCAERNRPQVYISLSAHSRWSVEKIVGVOLRG 298

QY 282 -----AELSRHHHHH 292
b 299 RAQGGGALRAPSQSGGAARSSRRHHHHH 327

RESULT 7
US-09-386-642-14
Sequence 14, Application US/09386642
Patent No. 6420157
GENERAL INFORMATION:
APPLICANT: Darrow, Andrew
APPLICANT: Qi, Jensen
APPLICANT: Andrade-Gordon, Patricia
TITLE OF INVENTION: Zymogen Activation System
FILE REFERENCE: ORT-1028
CURRENT APPLICATION NUMBER: US/09/386,642

CURRENT FILING DATE: 1999-08-31

NUMBER OF SEQ ID NOS: 60

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 14

LENGTH: 289

TYPE: PRT

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: Description of Artificial Sequence: Fusion gene

OTHER INFORMATION: with homo sapien serine protease catalytic domain

S-09-386-642-14

Query Match

Best Local Similarity 42.0%; Score 672.5; DB 4; Length 289;

Matches 148; Conservative 39; Mismatches 89; Indels 29; Gaps 11;

Y

1 MDSKGSQKSLRLLLLVSNLLLCQGVVSDYKDDDDVDAALAAPFDDDDKIVGGY-ALD 59

b

1 MDSKGSQKSLRLLLLVSNLLLCQGVVSDYKDDDDVDAALAAPFDDDDKIVGGY-NCLE 60

Y

60 VDSWPMQVSIQYDKQHVCGGSLDPHVLTAHCFRK-----HTDVFNWKVRAGSDK- 111

b

61 KHSQFWQAALFEKTRLLCGATLIAPRMLLTAHCLKRYIIVHLGQHNLOKEEGCEQRTA 120

Y

112 LGSPSLAVAKIIIBFNPMY-P-KD--NDIALMKLQPLTFSGTVRPICLPFPDEELTPA 168

b

121 TESFPHG-----FNNLFDKDRNDIMLVNAGSPVSTWAVRPLTUS--SRCVTAG 170

Y

169 TPLWIIWGFTKQNGKMSDILLQASVQVIDSTRCNADDAVQGBVTEKMMKAGIPEGGVD 228

b

171 TSLICSWGSTSSPQLRPLPHTLRCANITIEHQK--ENAYFGNITITMVCASVQEGKD 228

Y

229 TCQDGGPLMYQSDQHVHVIVSWGYG-CGGRSTPGVYTKVSAYLNIWVNWKAELSRH 287

b

229 SCQDGGPLVNCOS---LOGIISWGQDPCAITRKPGVYTKVCKYVDWIOETMKN-SRH 284

Y

288 HHHHH 292

b

285 HHHHH 289

RESULT 8

S-09-386-642-13

Sequence 13, Application US/09386642

Patent No. 6420157

GENERAL INFORMATION:

APPLICANT: Darrow, Andrew

APPLICANT: Qi, Jensen

APPLICANT: Andrade-Gordon, Patricia

TITLE OF INVENTION: Zymogen Activation System

FILE REFERENCE: ORT-1028

CURRENT APPLICATION NUMBER: US/09/386.642

CURRENT FILING DATE: 1999-08-31

NUMBER OF SEQ ID NOS: 60

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 13

LENGTH: 288

TYPE: PRT

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: Description of Artificial Sequence: Fusion gene

OTHER INFORMATION: with homo sapien serine protease catalytic domain

S-09-386-642-13

Query Match

Best Local Similarity 40.3%; Score 645; DB 4; Length 288;

Matches 146; Conservative 38; Mismatches 91; Indels 30; Gaps 11;

Y

1 MDSKGSQKSLRLLLLVSNLLLCQGVVSDYKDDDDVDAALAAPFDDDDKIVGGY-ALD 59

b

1 MDSKGSQKSLRLLLLVSNLLLCQGVVSDYKDDDDVDAALAAPFDDDDKIVGGY-NCLE 60

Y

60 VDSWPMQVSIQYDKQHVCGGSLDPHVLTAHCFR-KHTDVFNWKVRAGSDKJ-----GS 114

Db 61 PHSQFWQAALFOGQQLLCGVLVGNWLTAAHCKKPKYT-----VRLGDHSLQNKDGP 114

QY 115 FPSLAVAKIIIBFNPMY-----PKNDIALMKLQPLTFSGTVRPICLPFPDEELTPA 168

Db 115 EQEIPVQSIP---HPCYNSSDVEDHNDLMLQLRQASLGSKVKPISLA--DHCTQPG 169

QY 169 TPLWIIWGFTKQNGKMSDILLQASVQVIDSTRCNADDAVQGBVTEKMMKAGIPEGGVD 228

Db 170 QKTVSGWGTVPSPRENFDTLNCAEVKIPQKCC--EDAYPGQITDGMVCAGSSK-GAD 226

QY 229 TCQDGGPLMYQSDQHVHVIVSWGYG-CGGRSTPGVYTKVSAYLNIWVNWKAELSRH 287

Db 227 TCQDGGPLVNCG---ALQGITSWGSDPCGRDKPGVYTNICRYLDWIKKIIGSKSRH 283

QY 288 HHHHH 292

Db 284 HHHHH 288

RESULT 9

US-09-386-642-53

Sequence 53, Application US/09386642

Patent No. 6420157

GENERAL INFORMATION:

APPLICANT: Darrow, Andrew

APPLICANT: Qi, Jensen

APPLICANT: Andrade-Gordon, Patricia

TITLE OF INVENTION: Zymogen Activation System

FILE REFERENCE: ORT-1028

CURRENT APPLICATION NUMBER: US/09/386.642

CURRENT FILING DATE: 1999-08-31

NUMBER OF SEQ ID NOS: 60

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 53

LENGTH: 306

TYPE: PRT

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: Description of Artificial Sequence: Fusion gene of

OTHER INFORMATION: human protease F in CPEK2 zymogen vector

US-09-386-642-53

Query Match

Best Local Similarity 37.6%; Score 601.5; DB 4; Length 306;

Matches 126; Conservative 46; Mismatches 85; Indels 37; Gaps 10;

QY 28 VSDYKDDDDVDAALAAPFDDDDKIYGVYALDVDSWPMQVSIQYDKQHVCGGSLDPHVV 87

Db 21 VPDYKDDDD--DAAALAAPFDDDDKIYGVYALELGRWPWQGSLLRLWDSHVCGVLLSHRWA 78

QY 88 LTAHCFRKHDTDFN---WKVRAGSKLGSFSLAVAK-----IIITFENMYPKDN- 136

Db 79 LTAHCFRKHDTDFN---WKVRAGSKLGSFSLAVAK-----IIITFENMYPKDN- 136

QY 137 -DIALMKLQPLTFSGTVRPICLPFPDEELTPATLWIIWGFTKQNGKMSDILLQ-AS 194

Db 137 YDIALVLSAPVYTKHIQIPLQASTFEFENRTDCWTVGWYKEDALPSHTTQEVQ 196

QY 195 VQVIDSTRCN---ADDAVQGBVTEKMMKAGIPEGGVDTCQDGGPL-MYQSDQHVHVG 250

Db 197 VALINNSMCHLFLKYSFRKDFGDMVCAAGACGKADCFGDSGGPLACNGLWYQIGV 256

QY 251 VSMGYCGGSPSTPGVYTKVSAYLNIWV-----MKAELSRHHHHH 292

Db 257 VSMGYCGGSPSTPGVYTKVSAYLNIWV-----MKAELSRHHHHH 306

RESULT 10

US-09-386-642-12

Sequence 12, Application US/09386642

Patent No. 6420157

GENERAL INFORMATION:

APPLICANT: Darrow, Andrew
 APPLICANT: Qi, Jensen
 APPLICANT: Andrade-Gordon, Patricia
 TITLE OF INVENTION: Zymogen Activation System
 FILE REFERENCE: ORT-1028
 CURRENT APPLICATION NUMBER: US/09/386,642
 CURRENT FILING DATE: 1999-08-31
 NUMBER OF SEQ ID NOS: 60
 SOFTWARE: PatentIn Ver. 2.0
 SEQ ID NO 12
 LENGTH: 319
 TYPE: PRT
 ORGANISM: Artificial Sequence
 FEATURE:
 OTHER INFORMATION: Description of Artificial Sequence: Fusion gene
 OTHER INFORMATION: with homo sapien serine protease catalytic domain
 S-09-386-642-12
 Query Match 37.1%; Score 593; DB 4; Length 319;
 Best Local Similarity 44.2%; Pred. No. 1.2e-55;
 Matches 133; Conservative 35; Mismatches 95; Indels 38; Gaps 9;

Y 28 VSDYKDDDDVDAALAAPDDDDDKIVGGYALDVSWPMQVSIQYDKQHVCGGSILDPHWV 87
 b 21 VPDYKDD-DAALAAPDDDDDKIVGGYALDVSWPMQVSIQYDKQHVCGGSILDPHWV 78
 Y 88 LTAACHFRKHTDVFNWVKVRAGSKLGSPSLA-VAKIIIEFNPMYPKD---NDIALMKL 143
 b 79 LSAACHFSEHKEAYEVKLGKQDLSDSEDAKSTLKDIIIPHSYLOEGSQGDIALQL 138
 Y 144 QPLTFSTGTRVPCIPFDELTATPWLIIWGTQKNGKMS-DILLQASVOVIDSTR 202
 b 139 SRPITFSYRIPCIPFDELTATPWLIIWGTQKNGKMS-DILLQASVOVIDSTR 198
 Y 203 CNA---DDAYOCE---VTEKMMKAGIPGGVDTCCGDSGGPLMYQSD-OWHVGIVSGY 255
 b 199 CMLNLIADKPEPHFVQEDVVCAGYVGGKDACQDSGGPLSCPVEGLWLTGIVSGD 258
 Y 256 GCGSPSTPGVYTKVSAYNLIWY-----NWKAEI---SRHHHHH 291
 b 259 AGCARNPFGVYTLASSYASWISQVTELPQVVPQTSQPSDNLGSHLAFSSRRHHHH 318
 Y 292 H 292
 b 319 H 319

RESULT 11
 S-09-386-642-54
 Sequence 54, Application US/09386642
 Patent No. 6420157
 GENERAL INFORMATION:
 APPLICANT: Darrow, Andrew
 APPLICANT: Qi, Jensen
 APPLICANT: Andrade-Gordon, Patricia
 TITLE OF INVENTION: Zymogen Activation System
 FILE REFERENCE: ORT-1028
 CURRENT APPLICATION NUMBER: US/09/386,642
 CURRENT FILING DATE: 1999-08-31
 NUMBER OF SEQ ID NOS: 60
 SOFTWARE: PatentIn Ver. 2.0
 SEQ ID NO 54
 LENGTH: 284
 TYPE: PRT
 ORGANISM: Artificial Sequence
 FEATURE:
 OTHER INFORMATION: Description of Artificial Sequence: Human MH2
 OTHER INFORMATION: protease in PFEK zymogen vector
 S-09-386-642-54

Query Match 35.9%; Score 574; DB 4; Length 284;
 Best Local Similarity 44.2%; Pred. No. 1.1e-53;
 Matches 133; Conservative 39; Mismatches 103; Indels 26; Gaps 11;

QY 1 MDSKSSOKSRLLLLLVSNLLCOGVVSDYKDDDDVDAALAAPDDDDDKIVGGY-ALD 59
 Db 1 MDSKSSOKSRLLLLLVSNLLCOGVVSDYKDDDDVDAALAAPDDDDDKIVGGYNCLE 60
 QY 60 VDSWPQVSIQYDKQHVCGGSILDPHWVLTAAHCFRKHHTDVFNWVKVRAGSKL--GSPPS 117
 Db 61 PHSQPWQAALVMELELFCGVLVHPQWVLSAAHCFQN-----SYTIGLGLSHLEADQEPG 115
 QY 118 LAVAKIIIEFNPMYPKD---NDIALMKLQPLTSGTVRPICLPFFDEELTPATPLWII 174
 Db 116 SQKVEASLSVRHPEYNRPLANDMLIKLDESVSSEDTIRISIA--SQCTAGNSCLVS 173
 QY 175 GWGFTKQNGKMSDILLQASVOVIDSTRCNA--DDAYOCEVTEKMMKAGIPGGVDTCCG 232
 Db 174 GNGLLA--NGRMPVLQCVNVSVEEVCCKLYDPLXH---PSMFCAGGGHDKQDSCNG 227
 QY 233 DSGGPLMYOSDQWVHVGVISWGYG-CGGPSTPGVYTKVSAYNLIWYTNWKAEI-SRHHHHH 291
 Db 228 DSGGPLICNG--YLGGLVSGFKAPCGQGVGVYTNLCKFTWIEKTVQAS-SRHHHHH 283
 QY 292 H 292
 Db 284 H 284

RESULT 12
 US-08-508-448C-25
 ; Sequence 25, Application US/08508448C
 ; Patent No. 5804410
 ; GENERAL INFORMATION:
 ; APPLICANT: Kazuyoshi YAMAOKA et al.
 ; TITLE OF INVENTION: NUCLEIC ACID SEQUENCE ENCODING
 ; TITLE OF INVENTION: TRYPSIN-LIKE ENZYME AND PROCESS FOR PRODUCING THE ENZYME
 ; NUMBER OF SEQUENCES: 25
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Wenderoth, Lind & Bonack
 ; STREET: 805 Fifteenth Street, N.W., #700
 ; CITY: Washington
 ; STATE: D.C.
 ; COUNTRY: U.S.A.
 ; ZIP: 20005
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
 ; COMPUTER: IBM Compatible
 ; OPERATING SYSTEM: MS-DOS
 ; SOFTWARE: Wordperfect 5.1
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/508,448C
 ; FILING DATE: July 28, 1995
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER:
 ; FILING DATE:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Warren M. Cheek, Jr.
 ; REGISTRATION NUMBER: 33,367
 ; REFERENCE/DOCKET NUMBER:
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 202-371-8850
 ; TELEFAX:
 ; TELEX:
 ; INFORMATION FOR SEQ ID NO: 25:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 418 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; US-08-508-448C-25

Query Match 32.9%; Score 526.5; DB 1; Length 418;
 Best Local Similarity 41.8%; Pred. No. 2.5e-48;
 Matches 97; Conservative 49; Mismatches 79; Indels 7; Gaps 4;

b 302 PGSTAYVTGWAQOEYAGHTVPE-LRQGVRIISNDVCNHFHNGAILSCMLCAGVPQGG 360
Y 227 VDTCCGDSGGPLMYQSDQ--WHVVGIVSWGYCGGPSTPGVYTKVSAYLNWI 276
b 361 VDACCQDSGGPLVQEDSRRLWFIVGIVSWGDQCGLPKPGVYTRVTAYIDWI 412

earch completed: November 23, 2003, 07:53:47
ob time : 15.8611 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

M protein - protein search, using sw model

un on: November 23, 2003, 07:49:15 ; Search time 18.4759 Seconds
(without alignments)
2885.234 Million cell updates/sec

itle: US-09-607-745-9

erfect score: 1600

equences: 1 MDSGSSQKRLLLLVSN.....LNWIYVWKAHSRHHHHH 292

coring table:

Gapop 10.0 , Gapext 0.5

sarched: 666188 seqs, 182559486 residues

otal number of hits satisfying chosen parameters: 666188

inimum DB seq length: 0

aximum DB seq length: 2000000000

ost-processing: Minimum Match 0%

Listing first 45 summaries

atabase :

Published Applications AA:*
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/ACT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/PTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

result No.	Score	Query Match	Length	DB ID	Description
1	1266	79.1	406	9	US-09-851-588-6
2	1266	79.1	423	11	US-09-776-191-72
3	1266	79.1	423	15	US-10-264-820-23
4	1266	79.1	423	15	US-10-264-820-23
5	1266	79.1	432	11	US-09-888-257A-7
6	1266	79.1	432	11	US-09-946-374-275
7	1266	79.1	432	12	US-10-015-387A-275
8	1266	79.1	432	12	US-10-063-735-112
9	1266	79.1	432	12	US-10-006-130A-275
10	1266	79.1	432	12	US-10-199-672-330
11	1266	79.1	432	12	US-10-006-172A-275
12	1266	79.1	432	12	US-10-187-749-330
13	1266	79.1	432	12	US-10-194-457-330
14	1266	79.1	432	12	US-10-184-642-330
15	1266	79.1	432	12	US-10-196-747-330

16	1266	79.1	432	12	US-10-015-392A-275	Sequence 275, App
17	1266	79.1	432	12	US-10-017-253A-275	Sequence 275, App
18	1266	79.1	432	12	US-10-173-689-330	Sequence 330, App
19	1266	79.1	432	12	US-10-173-690-330	Sequence 330, App
20	1266	79.1	432	12	US-10-173-691-330	Sequence 330, App
21	1266	79.1	432	12	US-10-173-692-330	Sequence 330, App
22	1266	79.1	432	12	US-10-173-694-330	Sequence 330, App
23	1266	79.1	432	12	US-10-173-698-330	Sequence 330, App
24	1266	79.1	432	12	US-10-173-699-330	Sequence 330, App
25	1266	79.1	432	12	US-10-173-707-330	Sequence 330, App
26	1266	79.1	432	12	US-10-174-569-330	Sequence 330, App
27	1266	79.1	432	12	US-10-174-583-330	Sequence 330, App
28	1266	79.1	432	12	US-10-174-587-330	Sequence 330, App
29	1266	79.1	432	12	US-10-174-589-330	Sequence 330, App
30	1266	79.1	432	12	US-10-174-591-330	Sequence 330, App
31	1266	79.1	432	12	US-10-175-736-330	Sequence 330, App
32	1266	79.1	432	12	US-10-175-742-330	Sequence 330, App
33	1266	79.1	432	12	US-10-175-744-330	Sequence 330, App
34	1266	79.1	432	12	US-10-175-745-330	Sequence 330, App
35	1266	79.1	432	12	US-10-175-748-330	Sequence 330, App
36	1266	79.1	432	12	US-10-175-751-330	Sequence 330, App
37	1266	79.1	432	12	US-10-175-754-330	Sequence 330, App
38	1266	79.1	432	12	US-10-176-480-330	Sequence 330, App
39	1266	79.1	432	12	US-10-176-489-330	Sequence 330, App
40	1266	79.1	432	12	US-10-176-754-330	Sequence 330, App
41	1266	79.1	432	12	US-10-176-755-330	Sequence 330, App
42	1266	79.1	432	12	US-10-176-759-330	Sequence 330, App
43	1266	79.1	432	12	US-10-176-920-330	Sequence 330, App
44	1266	79.1	432	12	US-10-176-922-330	Sequence 330, App
45	1266	79.1	432	12	US-10-176-924-330	Sequence 330, App

ALIGNMENTS

RESULT 1

US-09-851-588-6
; Sequence 6, Application US/09851588
; Patent No. US20020042067A1
; GENERAL INFORMATION:
; APPLICANT: Mack, David
; APPLICANT: Wilson, Keith E.
; TITLE OF INVENTION: NOVEL METHODS OF DIAGNOSING COLORECTAL CANCER, COMPOSITIONS, AND A
; FILE OF INVENTION: OF SCREENING FOR COLORECTAL CANCER MODULATORS
; FILE REFERENCE: A-68829-1/DJB/JJD/AMS
; CURRENT APPLICATION NUMBER: US/09/851,588
; CURRENT FILING DATE: 2001-09-24
; PRIOR APPLICATION NUMBER: US 09/642,252
; PRIOR FILING DATE: 2000-08-17
; PRIOR APPLICATION NUMBER: US 09/656,002
; PRIOR FILING DATE: 2000-09-06
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
; LENGTH: 406
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-851-588-6

Query Match 79.1%; Score 1266; DB 9; Length 406;
Best Local Similarity 97.4%; Pred. No. 5.8e-117;
Matches 228; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
QY 51 KIYGGVALVDSPWQVSIQYDKQHVCGGSIIDPHWVLTAAHCFRKHDTVFNNKVRAGSD 110
Db 173 RVYGGEEASVDSWPWQVSIQYDKQHVCGGSIIDPHWVLTAAHCFRKHDTVFNNKVRAGSD 232
QY 111 KLGSPFLAVAKIIIIIEFNPMYKNDIALMKLOPILTFSGIVRPICLPFFDELPATP 170
Db 233 KLGSPFLAVAKIIIIIEFNPMYKNDIALMKLOPILTFSGIVRPICLPFFDELPATP 292
QY 171 LWTIGWFTKQNGKMSDILLQASVQVIDSTRCNADDAVQGEVTERKVMCAGIEGGVDTC 230

[illegible][illegible]

PRIOR FILING DATE: 1998-10-22
 PRIOR APPLICATION NUMBER: 60/105266
 PRIOR FILING DATE: 1998-10-22
 PRIOR APPLICATION NUMBER: 60/105693
 PRIOR FILING DATE: 1998-10-26
 PRIOR APPLICATION NUMBER: 60/105694
 PRIOR FILING DATE: 1998-10-26
 PRIOR APPLICATION NUMBER: 60/105807

Query Match	79.1%;	Score 1266;	DB 11;	Length 432;
Best Local Similarity	97.4%;	Prod. No. 6.3e-117;		
Matches 228;	Conservative 2;	Mismatches 4;	Indels 0;	Gaps 0;
Y	51	KIVGGVALVDSPWQVSIQYDKQHVCGGSIIDPHWVLTAAHCFRKHHDVENWVKVRGSD	110	
b	199	RVVGGEEASVDSWPQVSIQYDKQHVCGGSIIDPHWVLTAAHCFRKHHDVENWVKVRGSD	258	
Y	111	KLGSFSLAVAKIIIEFNPMPKONDIALMKLOPLTFSGTVRPICLPFFDEELTPATP	170	
b	259	KLGSFSLAVAKIIIEFNPMPKONDIALMKLOPLTFSGTVRPICLPFFDEELTPATP	318	
Y	171	LWTIIGWFTKGGRKMSDILLQASVQVIDSTRCNADDAVQGEVTEKMKACIGPEGGVDTCC	230	
b	319	LWTIIGWFTKGGRKMSDILLQASVQVIDSTRCNADDAVQGEVTEKMKACIGPEGGVDTCC	378	
Y	231	QDSDSGPLMYQSDQVHVGIVSWGVCGCGPSPPGVYTKVSAYLNIYNVWKAEI	284	
b	379	QDSDSGPLMYQSDQVHVGIVSWGVCGCGPSPPGVYTKVSAYLNIYNVWKAEI	432	

ESULT 7
S-10-015-387A-275
Sequence 275, Application US/10015387A

Publication No. US20030135034A1
 GENERAL INFORMATION:
 APPLICANT: Baker, Kevin P.
 APPLICANT: Botstein, David
 APPLICANT: Desnovers, Iuc
 APPLICANT: Eaton, Dan I.
 APPLICANT: Ferrara, Napoleone
 APPLICANT: Fong, Sherman
 APPLICANT: Gao, Wei-Qiang
 APPLICANT: Goddard, Audrey
 APPLICANT: Godowski, Paul J.
 APPLICANT: Grimaldi, Christopher J.
 APPLICANT: Gurney, Austin L.
 APPLICANT: Hillan, Kenneth J.
 APPLICANT: Pan, James
 APPLICANT: Paoni, Nicholas F.
 TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
 TITLE OF INVENTION: Acids Encoding the Same
 FILE REFERENCE: P2830PlC54
 CURRENT APPLICATION NUMBER: US/10/015,387A
 CURRENT FILING DATE: 2001-12-12
 Prior Application removed - See File Wrapper or Palm
 NUMBER OF SEQ ID NOS: 477
 SEQ ID NO 275
 LENGTH: 432

ORGANISM: HOMO sapiens									
S-10-015-387A-275									
Query Match									
79.1%; Score 1266; DB 12; Length 432;									
Best Local Similarity 97.4%; Pred. No. 6.3e-117;									
Matches 228; Conservative 2; Mismatches 4; Indels 0; Gaps 0;									
y	51	KIVGVALDVDS	PNWQVSIQYDKQHYCGGS	ILDPHWLTAACFCRKHTDVFNW	KVRAGSD	110			
b	199	RVYGGBEAS	VDSPWQVSIQYDKQHYCGGS	ILDPHWLTAACFCRKHTDVFNW	KVRAGSD	258			
		:	:	:	:				
y	111	KLGSFFSLAVAK	IIIEFNPMPKNDIA	LMKLOPLTFSGTVR	ICLPFFDEELTPATP	170			
b	259	KLGSFFSLAVAK	IIIEFNPMPKNDIA	LMKLOPLTFSGTVR	ICLPFFDEELTPATP	318			

Qy	171	LWIIIGWFTKQNGKMSDILLLOASVOVIDSTRCNADDAIQSEVTEKMNKAGIPEGGVDT	230
Db	319	LWIIIGWFTKQNGKMSDILLLOASVOVIDSTRCNADDAIQSEVTEKMNKAGIPEGGVDT	378
Qy	231	QSDSGGFLMYTQSDQHWVVGIVSWGVCGGGSPGTYTKVLSAYLNIWYINVKAE	284
Db	379	QSDSGGFLMYTQSDQHWVVGIVSWGVCGGGSPGTYTKVLSAYLNIWYINVKAE	432

```

RESULT 8
US-10-063-735-112
; Sequence 112, Application US/10063735
; Publication No. US2003013882A1
; GENERAL INFORMATION:
; APPLICANT: Eaton, Dan L.
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3230R1C1
; CURRENT APPLICATION NUMBER: US/10/063,735
; CURRENT FILING DATE: 2002-05-08

```

/ COMMENT: LANDING DATE: 2002-03-08
 / Prior Application removed - See Palm or File Wrapper
 / NUMBER OF SEQ ID NOS: 170

```

; SEQ ID NO 112
; LENGTH: 432
; TYPE: PRT
; ORGANISM: H
US-10-063-735-1

```

Query Match	79.1%;	Score 1266;	DB 12;	Length 432;
Best Local Similarity	97.4%;	Pred. No. 6.3e-117;		
Matches 228;	Conservative 2;	Mismatches 4;	Indels 0;	Gaps 0;
QY	51	KIVGGVALDVDSWPQVSVIQYDKOHVCGGSIILDPHWLITAAHCFRKHTDVFNVKVRGSD	110	
Db	199	RVVGGEASVDSWPQVSVIQYDQHVCGGSIILDPHWLITAAHCFRKHTDVFNVKVRGSD	258	
QY	111	KIGSFPSLAVAKIIITFENPMYKDNIDIALMKLOFLTFSGTVRPICLPFPDEELITPATP	170	
Db	259	KIGSFPSLAVAKIIITFENPMYKDNIDIALMKLOFLTFSGTVRPICLPFPDEELITPATP	318	
QY	171	LMIIIGWGFTKQNGKMSDILLQASVOVIDSTRCNADDAYQGEVTEKMCAGIPGGVDTC	230	
Db	319	LMIIIGWGFTKQNGKMSDILLQASVOVIDSTRCNADDAYQGEVTEKMCAGIPGGVDTC	378	
QY	231	QGDSGGLMYQSDQHVHVGIVSWGYCGGFPSTPGVYTKVSAYLNWIVNVWKAEL	284	
Db	379	QGDSGGLMYQSDQHVHVGIVSWGYCGGFPSTPGVYTKVSAYLNWIVNVWKAEL	432	

RESULT 9
US-10-006-130A-275
/ Sequence 275, Application US10006130A
/ Publication No. US20030148375A1
/ GENERAL INFORMATION:
/ APPLICANT: Baker, Kevin P.
/ APPLICANT: Botstein, David
/ APPLICANT: Desnoyers, Luc
/ APPLICANT: Eaton, Dan I.
/ APPLICANT: Ferrara, Napoleone
/ APPLICANT: Fong, Sherman
/ APPLICANT: Gao, Wei-qiang
/ APPLICANT: Goddard, Audrey
/ APPLICANT: Godowski, Paul J.
/ APPLICANT: Grimaldi, Christopher J.

```
; PRIOR APPLICATION NUMBER: 60/063540
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/063541
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/063542
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/063543
; PRIOR FILING DATE: 1997-10-28
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 330
; LENGTH: 432
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-199-672-330

Query Match          79.1%; Score 1266; DB 12; Length 432;
Best Local Similarity 97.4%; Pred. No. 6.3e-117;
Matches 228; Conservative 2; Mismatches 4; Indels 0; Gaps

QY      51 KIVGGVALVDSPWQVSIQYDKQHVCGGSILDPHWLTAACFRKHHTDVFNNKVRAGSD
Db      199 RVVGGEASVDSWPWQVSIQYDKQHVCGGSILDPHWLTAACFRKHHTDVFNNKVRAGSD
        :|||
QY      111 KLGSFPLAVAKIIIEFNPMYPKNDIALMKLOPFLFSFGTVRPICLPPFFDEELTPATP
Db      259 KLGSFPLAVAKIIIEFNPMYPKNDIALMKLOPFLFSFGTVRPICLPPFFDEELTPATP
        :|||
QY      171 LMTIGWFTKQNGKMSDILLQASVQVIDSTRCNADDAYQGSEVTETKMCCAGIPGGGVDFTC
Db      319 LMTIGWFTKQNGKMSDILLQASVQVIDSTRCNADDAYQGSEVTETKMCCAGIPGGGVDFTC
        :|||
QY      231 QGDSSGLMYQSDQWHVGVISWGYGCGGSPPTGYTKVSALNINYNWKAEI 284
Db      379 QGDSSGLMYQSDQWHVGVISWGYGCGGSPPTGYTKVSALNINYNWKAEI 432

RESULT 11
US-10-006-172A-275
; Sequence 275, Application US/10006172A
; Publication No. US2003015300A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan I.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth J.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; TITLE OF INVENTION: Acids Encoding the Same
; FILE REFERENCE: P2830PIC11
; CURRENT APPLICATION NUMBER: US/10/006,172A
; CURRENT FILING DATE: 2002-03-19
; PRIOR APPLICATION NUMBER: 60/098716
; PRIOR FILING DATE: 1998-09-01
; PRIOR APPLICATION NUMBER: 60/098723
; PRIOR FILING DATE: 1998-09-01
; PRIOR APPLICATION NUMBER: 60/098749
; PRIOR FILING DATE: 1998-09-01
; PRIOR APPLICATION NUMBER: 60/098750
; PRIOR FILING DATE: 1998-09-01
; PRIOR APPLICATION NUMBER: 60/098803
; PRIOR FILING DATE: 1998-09-02
; PRIOR APPLICATION NUMBER: 60/098821
; PRIOR FILING DATE: 1998-09-02
; PRIOR APPLICATION NUMBER: 60/098843
; PRIOR FILING DATE: 1998-09-02
```

7	PRIOR FILING DATE: 1998-09-23	
7	PRIOR APPLICATION NUMBER: 60/101476	
7	PRIOR FILING DATE: 1998-09-23	
7	PRIOR APPLICATION NUMBER: 60/101477	
7	PRIOR FILING DATE: 1998-09-23	
7	PRIOR APPLICATION NUMBER: 60/101479	
7	PRIOR FILING DATE: 1998-09-23	
7	PRIOR APPLICATION NUMBER: 60/101738	
7	PRIOR FILING DATE: 1998-09-24	
7	PRIOR APPLICATION NUMBER: 60/101741	
7	PRIOR FILING DATE: 1998-09-24	
7	PRIOR APPLICATION NUMBER: 60/101743	
7	PRIOR FILING DATE: 1998-09-24	
7	PRIOR APPLICATION NUMBER: 60/101915	
7	PRIOR FILING DATE: 1998-09-24	
7	PRIOR APPLICATION NUMBER: 60/101916	
7	PRIOR FILING DATE: 1998-09-24	
7	PRIOR APPLICATION NUMBER: 60/102207	
7	PRIOR FILING DATE: 1998-09-29	
7	PRIOR APPLICATION NUMBER: 60/102240	
7	PRIOR FILING DATE: 1998-09-29	
7	PRIOR APPLICATION NUMBER: 60/102307	
7	PRIOR FILING DATE: 1998-09-29	
7	PRIOR APPLICATION NUMBER: 60/102330	
7	PRIOR FILING DATE: 1998-09-29	
7	PRIOR APPLICATION NUMBER: 60/102331	
7	PRIOR FILING DATE: 1998-09-29	
7	PRIOR APPLICATION NUMBER: 60/102484	
7	PRIOR FILING DATE: 1998-09-30	
7	PRIOR APPLICATION NUMBER: 60/102487	
7	PRIOR FILING DATE: 1998-09-30	
7	PRIOR APPLICATION NUMBER: 60/102570	
7	PRIOR FILING DATE: 1998-09-30	
7	PRIOR APPLICATION NUMBER: 60/102571	
7	PRIOR FILING DATE: 1998-09-30	
7	PRIOR APPLICATION NUMBER: 60/102584	
7	PRIOR FILING DATE: 1998-10-01	
7	PRIOR APPLICATION NUMBER: 60/102687	
7	PRIOR FILING DATE: 1998-10-01	
7	PRIOR APPLICATION NUMBER: 60/102965	
7	PRIOR FILING DATE: 1998-10-02	
7	PRIOR APPLICATION NUMBER: 60/103258	
7	PRIOR FILING DATE: 1998-10-06	
7	PRIOR APPLICATION NUMBER: 60/103314	
7	PRIOR FILING DATE: 1998-10-07	
7	PRIOR APPLICATION NUMBER: 60/103315	
7	PRIOR FILING DATE: 1998-10-07	
7	PRIOR APPLICATION NUMBER: 60/103328	
7	PRIOR FILING DATE: 1998-10-07	
7	PRIOR APPLICATION NUMBER: 60/103395	
7	PRIOR FILING DATE: 1998-10-07	
7	PRIOR APPLICATION NUMBER: 60/103396	
7	PRIOR FILING DATE: 1998-10-07	
7	PRIOR APPLICATION NUMBER: 60/103401	
7	PRIOR FILING DATE: 1998-10-07	
7	PRIOR APPLICATION NUMBER: 60/103449	
7	PRIOR FILING DATE: 1998-10-06	
7	PRIOR APPLICATION NUMBER: 60/103633	
7	PRIOR FILING DATE: 1998-10-08	
7	PRIOR APPLICATION NUMBER: 60/103678	
7	PRIOR FILING DATE: 1998-10-08	
7	PRIOR APPLICATION NUMBER: 60/103679	
7	PRIOR FILING DATE: 1998-10-08	
7	PRIOR APPLICATION NUMBER: 60/103711	
7	PRIOR FILING DATE: 1998-10-08	
7	PRIOR APPLICATION NUMBER: 60/104257	
7	PRIOR FILING DATE: 1998-10-14	
7	PRIOR APPLICATION NUMBER: 60/104987	
7	PRIOR FILING DATE: 1998-10-20	
7	PRIOR APPLICATION NUMBER: 60/105000	
7	PRIOR FILING DATE: 1998-10-20	
7	PRIOR APPLICATION NUMBER: 60/105002	
7	PRIOR FILING DATE: 1998-10-20	

PRIOR APPLICATION NUMBER: 60/105104
PRIOR FILING DATE: 1998-10-21
PRIOR APPLICATION NUMBER: 60/105169
PRIOR FILING DATE: 1998-10-22
PRIOR APPLICATION NUMBER: 60/105266
PRIOR FILING DATE: 1998-10-22
PRIOR APPLICATION NUMBER: 60/105693
PRIOR FILING DATE: 1998-10-26
PRIOR APPLICATION NUMBER: 60/105694
PRIOR FILING DATE: 1998-10-26
PRIOR APPLICATION NUMBER: 60/105807
PRIOR FILING DATE: 1998-10-27
PRIOR APPLICATION NUMBER: 60/105881
PRIOR FILING DATE: 1998-10-27
PRIOR APPLICATION NUMBER: 60/105882
PRIOR FILING DATE: 1998-10-27
PRIOR APPLICATION NUMBER: 60/106023
PRIOR FILING DATE: 1998-10-28

Query Match 79.1%; Score 1266; DB 12; Length 432;
Best Local Similarity 97.4%; Pred. No. 6.3e-117;
Matches 228; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
Y 51 KIVGGYALDVDSWPQVSTQYDKQHVCGSILDPHWLTAACHCFKHTDVFNNKVRAGSD 110
b 199 RVVGGEEASVDSWPQVSTQYDKQHVCGSILDPHWLTAACHCFKHTDVFNNKVRAGSD 258
Y 111 KLGSPFSLAVAKIIIEFNPMYPKNDIALMKLOPLTFSGTVRPICLPFFDEELTPATP 170
b 259 KLGSPFSLAVAKIIIEFNPMYPKNDIALMKLOPLTFSGTVRPICLPFFDEELTPATP 318
Y 171 LWIIGWGTQKNGKMSDILLQASVQVIDSTRCNADDAVQGEVTEKMKCAGIPEGGVDT 230
b 319 LWIIGWGTQKNGKMSDILLQASVQVIDSTRCNADDAVQGEVTEKMKCAGIPEGGVDT 378
Y 231 QDGGGGLMYQSDQHVHVGVSVGWGCGGPGSTPGVYTKVSAYLNWYNNWKAEL 284
b 379 QDGGGGLMYQSDQHVHVGVSVGWGCGGPGSTPGVYTKVSAYLNWYNNWKAEL 432

RESULT 12
US-10-187-749-330
Sequence 330, Application US/10187749
Publication No. US20030153036A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Chen, Jian
APPLICANT: Desnoyers, Luc
APPLICANT: Goddard, Audrey
APPLICANT: Goddard, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Pan, James
APPLICANT: Smith, Victoria
APPLICANT: Watanabe, Colin K.
APPLICANT: Wood, William I.
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P3430R1C1
CURRENT APPLICATION NUMBER: US/10187,749
CURRENT FILING DATE: 2002-07-01
PRIOR APPLICATION NUMBER: US/10/052,586
PRIOR FILING DATE: 2002-01-15
PRIOR APPLICATION NUMBER: 60/059263
PRIOR FILING DATE: 1997-09-18
PRIOR APPLICATION NUMBER: 60/059266
PRIOR FILING DATE: 1997-09-18
PRIOR APPLICATION NUMBER: 60/062250
PRIOR FILING DATE: 1997-10-17
PRIOR APPLICATION NUMBER: 60/063120
PRIOR FILING DATE: 1997-10-24
PRIOR APPLICATION NUMBER: 60/063121
PRIOR FILING DATE: 1997-10-24

PRIOR APPLICATION NUMBER: 60/063486
PRIOR FILING DATE: 1997-10-21
PRIOR APPLICATION NUMBER: 60/063540
PRIOR FILING DATE: 1997-10-28
PRIOR APPLICATION NUMBER: 60/063541
PRIOR FILING DATE: 1997-10-28
PRIOR APPLICATION NUMBER: 60/063544
PRIOR FILING DATE: 1997-10-28
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 612
SEQ ID NO 330
LENGTH: 432
TYPE: PRT
ORGANISM: Homo Sapien
US-10-187-749-330

Query Match 79.1%; Score 1266; DB 12; Length 432;
Best Local Similarity 97.4%; Pred. No. 6.3e-117;
Matches 228; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
QY 51 KIVGGYALDVDSWPQVSTQYDKQHVCGSILDPHWLTAACHCFKHTDVFNNKVRAGSD 110
Db 199 RVVGGEEASVDSWPQVSTQYDKQHVCGSILDPHWLTAACHCFKHTDVFNNKVRAGSD 258
QY 111 KLGSPFSLAVAKIIIEFNPMYPKNDIALMKLOPLTFSGTVRPICLPFFDEELTPATP 170
Db 259 KLGSPFSLAVAKIIIEFNPMYPKNDIALMKLOPLTFSGTVRPICLPFFDEELTPATP 318
QY 171 LWIIGWGTQKNGKMSDILLQASVQVIDSTRCNADDAVQGEVTEKMKCAGIPEGGVDT 230
Db 319 LWIIGWGTQKNGKMSDILLQASVQVIDSTRCNADDAVQGEVTEKMKCAGIPEGGVDT 378
QY 231 QDGGGGLMYQSDQHVHVGVSVGWGCGGPGSTPGVYTKVSAYLNWYNNWKAEL 284
Db 379 QDGGGGLMYQSDQHVHVGVSVGWGCGGPGSTPGVYTKVSAYLNWYNNWKAEL 432

RESULT 13
US-10-194-457-330
Sequence 330, Application US/10194457
Publication No. US20030153037A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Chen, Jian
APPLICANT: Desnoyers, Luc
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Pan, James
APPLICANT: Smith, Victoria
APPLICANT: Watanabe, Colin K.
APPLICANT: Wood, William I.
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P3430R1C296
CURRENT APPLICATION NUMBER: US/10/194,457
CURRENT FILING DATE: 2002-07-11
PRIOR APPLICATION NUMBER: 10/052586
PRIOR FILING DATE: 2002-01-15
PRIOR APPLICATION NUMBER: 60/059263
PRIOR FILING DATE: 1997-09-18
PRIOR APPLICATION NUMBER: 60/059266
PRIOR FILING DATE: 1997-09-18
PRIOR APPLICATION NUMBER: 60/062250
PRIOR FILING DATE: 1997-10-17
PRIOR APPLICATION NUMBER: 60/063120
PRIOR FILING DATE: 1997-10-24
PRIOR APPLICATION NUMBER: 60/063121
PRIOR FILING DATE: 1997-10-24
PRIOR APPLICATION NUMBER: 60/063486
PRIOR FILING DATE: 1997-10-21
PRIOR APPLICATION NUMBER: 60/063540

PRIOR FILING DATE: 1997-10-28
PRIOR APPLICATION NUMBER: 60/063541
PRIOR FILING DATE: 1997-10-28
PRIOR APPLICATION NUMBER: 60/063544
PRIOR FILING DATE: 1997-10-28
Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 612
SEQ ID NO 330
LENGTH: 432
TYPE: PRT
ORGANISM: Homo Sapien
S-10-194-457-330

Query Match 79.1%; Score 1266; DB 12; Length 432;
Best Local Similarity 97.4%; Pred. No. 6.3e-117;
Matches 228; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

```
Y 51 KIYGGYALVDSDWPNQVSIQYDKQHVCGGSIIDPHWVLTAAHCFRKHDTDFNWKVRAGSD 110
b 199 RVVGGEEASVDSWPNQVSIQYDKQHVCGGSIIDPHWVLTAAHCFRKHDTDFNWKVRAGSD 258
Y 111 KLGSPFLAVAKIIIEFNPMYPKNDIALMKLOPPLTFSGTVRPICLPFFDEELTPATP 170
b 259 KLGSPFLAVAKIIIEFNPMYPKNDIALMKLOPPLTFSGTVRPICLPFFDEELTPATP 318
Y 171 LWIIGWFTKQNGKMSDILLQASVQVIDSTRCNADDAYQGEVTEKMKCAGIPEGGYDTC 230
b 319 LWIIGWFTKQNGKMSDILLQASVQVIDSTRCNADDAYQGEVTEKMKCAGIPEGGYDTC 378
Y 231 QGDSGGPLMYQSDQWVHVGVISWGYGCGGSPSTPGVYTKVSAYLNWIYVWKAEL 284
b 379 QGDSGGPLMYQSDQWVHVGVISWGYGCGGSPSTPGVYTKVSAYLNWIYVWKAEL 432
```

ESULT 14

S-10-184-642-330

Sequence 330, Application US/10184642

Publication No. US20030157635A1

GENERAL INFORMATION:

APPLICANT: Baker, Kevin P.

APPLICANT: Chen, Jian

APPLICANT: Desnoyers, Luc

APPLICANT: Goddard, Audrey

APPLICANT: Godowski, Paul J.

APPLICANT: Gurney, Austin L.

APPLICANT: Pan, James

APPLICANT: Smith, Victoria

APPLICANT: Watanabe, Colin K.

APPLICANT: Wood, William I.

APPLICANT: Zhang, Zemin

TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC

FILE REFERENCE: P3430R1C194

CURRENT FILING DATE: 2002-06-27

Prior Application removed - See File Wrapper or Palm

NUMBER OF SEQ ID NOS: 612

SEQ ID NO 330

LENGTH: 432

TYPE: PRT

ORGANISM: Homo Sapien

S-10-194-642-330

Query Match 79.1%; Score 1266; DB 12; Length 432;
Best Local Similarity 97.4%; Pred. No. 6.3e-117;
Matches 228; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

```
Y 51 KIYGGYALVDSDWPNQVSIQYDKQHVCGGSIIDPHWVLTAAHCFRKHDTDFNWKVRAGSD 110
b 199 RVVGGEEASVDSWPNQVSIQYDKQHVCGGSIIDPHWVLTAAHCFRKHDTDFNWKVRAGSD 258
Y 111 KLGSPFLAVAKIIIEFNPMYPKNDIALMKLOPPLTFSGTVRPICLPFFDEELTPATP 170
```

```
Db 259 KLGSPFLAVAKIIIEFNPMYPKNDIALMKLOPPLTFSGTVRPICLPFFDEELTPATP 318
Qy 171 LWIIGWFTKQNGKMSDILLQASVQVIDSTRCNADDAYQGEVTEKMKCAGIPEGGYDTC 230
Db 319 LWIIGWFTKQNGKMSDILLQASVQVIDSTRCNADDAYQGEVTEKMKCAGIPEGGYDTC 378
Qy 231 QGDSGGPLMYQSDQWVHVGVISWGYGCGGSPSTPGVYTKVSAYLNWIYVWKAEL 284
Db 379 QGDSGGPLMYQSDQWVHVGVISWGYGCGGSPSTPGVYTKVSAYLNWIYVWKAEL 432
```

RESULT 15

US-10-196-747-330

Sequence 330, Application US/10196747

Publication No. US20030162250A1

GENERAL INFORMATION:

APPLICANT: Baker, Kevin P.

APPLICANT: Chen, Jian

APPLICANT: Desnoyers, Luc

APPLICANT: Goddard, Audrey

APPLICANT: Godowski, Paul J.

APPLICANT: Gurney, Austin L.

APPLICANT: Pan, James

APPLICANT: Smith, Victoria

APPLICANT: Watanabe, Colin K.

APPLICANT: Wood, William I.

APPLICANT: Zhang, Zemin

TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC

FILE REFERENCE: P3430R1C346

CURRENT APPLICATION NUMBER: US/10/196,747

CURRENT FILING DATE: 2002-07-16

Prior Application removed - See File Wrapper or Palm

NUMBER OF SEQ ID NOS: 612

SEQ ID NO 330

LENGTH: 432

TYPE: PRT

ORGANISM: Homo Sapien

US-10-196-747-330

Query Match 79.1%; Score 1266; DB 12; Length 432;
Best Local Similarity 97.4%; Pred. No. 6.3e-117;
Matches 228; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

```
Qy 51 KIYGGYALVDSDWPNQVSIQYDKQHVCGGSIIDPHWVLTAAHCFRKHDTDFNWKVRAGSD 110
Db 199 RVVGGEEASVDSWPNQVSIQYDKQHVCGGSIIDPHWVLTAAHCFRKHDTDFNWKVRAGSD 258
Qy 111 KLGSPFLAVAKIIIEFNPMYPKNDIALMKLOPPLTFSGTVRPICLPFFDEELTPATP 170
Db 259 KLGSPFLAVAKIIIEFNPMYPKNDIALMKLOPPLTFSGTVRPICLPFFDEELTPATP 318
Qy 171 LWIIGWFTKQNGKMSDILLQASVQVIDSTRCNADDAYQGEVTEKMKCAGIPEGGYDTC 230
Db 319 LWIIGWFTKQNGKMSDILLQASVQVIDSTRCNADDAYQGEVTEKMKCAGIPEGGYDTC 378
Qy 231 QGDSGGPLMYQSDQWVHVGVISWGYGCGGSPSTPGVYTKVSAYLNWIYVWKAEL 284
Db 379 QGDSGGPLMYQSDQWVHVGVISWGYGCGGSPSTPGVYTKVSAYLNWIYVWKAEL 432
```

Search completed: November 23, 2003, 07:54:48

Job time : 19.4759 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

M protein - protein search, using sw model

run on: November 23, 2003, 07:45:50 ; Search time 19.2792 Seconds
(without alignments)
1456.555 Million cell updates/sec

title: US-09-607-745-9

effect score: 1600

sequence: 1 MDSKGSSQKGRLLLLLVSN.....LNWIYNVWKABLSRHHHHH 292

coring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

searched: 283308 seqs, 96168682 residues

total number of hits satisfying chosen parameters: 283308

minimum DB seq length: 0

maximum DB seq length: 2000000000

post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

atabase :

PIR 76.*

1: Pirl.*

2: Pirl2.*

3: Pirl3.*

4: Pirl4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

result No.	Score	Query Match	Length	DB ID	Description
1	505.5	31.6	417	1 S00845	hepsin (EC 3.4.21.)
2	499	31.2	436	2 JX0172	acrosin (EC 3.4.21)
3	493	30.8	418	2 A37344	acrosin (EC 3.4.21)
4	492	30.8	638	1 K0MSPL	plasma kallikrein
5	491	30.7	437	2 S18407	acrosin (EC 3.4.21)
6	489	30.6	638	1 K0RTPL	plasma kallikrein
7	479.5	30.0	416	1 S33777	hepsin (EC 3.4.21)
8	479	29.9	638	1 K0HUP	plasma kallikrein
9	478.5	29.9	421	2 S29599	acrosin (EC 3.4.21)
10	476	29.8	421	1 S11674	acrosin (EC 3.4.21)
11	473.5	29.6	415	1 A34170	acrosin (EC 3.4.21)
12	470.5	29.4	246	2 B25528	acrosin (EC 3.4.21)
13	470	29.4	1524	2 T30337	polypyrroline - Afri
14	469	29.3	277	2 S35340	trypsin (EC 3.4.21)
15	464.5	29.0	247	1 B25852	trypsin (EC 3.4.21)
16	463.5	29.0	246	1 TRRT1	trypsin (EC 3.4.21)
17	463.5	29.0	267	2 S40006	trypsin (EC 3.4.21)
18	463.5	29.0	431	2 S47538	acrosin (EC 3.4.21)
19	459.5	28.7	247	1 TRDG	trypsin (EC 3.4.21)
20	456.5	28.5	275	2 S40007	trypsin (EC 3.4.21)
21	451.5	28.2	246	1 TRRT2	trypsin (EC 3.4.21)
22	450	28.1	274	2 S35339	trypsin (EC 3.4.21)
23	449.5	28.1	247	1 A25852	trypsin (EC 3.4.21)
24	444	27.8	304	2 S33496	trypsin (EC 3.4.21)
25	443.5	27.7	247	2 S13813	trypsin (EC 3.4.21)
26	443	27.7	420	2 A55283	acrosin (EC 3.4.21)
27	443	27.7	790	1 PLPG	plasmin (EC 3.4.21)
28	442.5	27.7	231	1 TRFGTR	trypsin (EC 3.4.21)
29	442.5	27.7	343	1 A57014	proctasin (EC 3.4.

coagulation factor
plasmin (EC 3.4.21)
trypsin-related pr
membrane-bound arg
chymotrypsin (EC 3
plasmin (EC 3.4.21)
trypsin (EC 3.4.21)
trypsin (EC 3.4.21)
enteropeptidase (E
enteropeptidase (E
trypsin (EC 3.4.21)
trypsin (EC 3.4.21)
plasmin (EC 3.4.21)
enteropeptidase (E
plasmin (EC 3.4.21)
trypsin-like prote

ALIGNMENTS

RESULT 1

S00845

hepsin (EC 3.4.21.-) - human

C/Species: Homo sapiens (man)

C/Date: 31-Dec-1988 #sequence_revision 31-Dec-1988 #text_change 18-Jun-1999

C/Accession: S00845

R/Leytus, S.P.; Loeb, K.R.; Hagen, F.S.; Kurachi, K.; Davie, E.W.

Biochemistry 27, 1067-1074, 1988

A/Title: A novel trypsin-like serine protease (hepsin) with a putative transmembrane domain

A/Reference number: S00845; MUID:88209431; PMID:2835076

A/Accession: S00845

A/Molecule type: mRNA

A/Residues: 1-417 <LEV>

A/Cross-references: EMBL:X07732; NID:g32063; PIDN:CAA30558.1; PID:g32064

C/Genetics:

A/Gene: GDB:HPN; TMRSS1; hepsin

A/Cross-references: GDB:I35685; OMIM:142440

A/Map position: 19q11-19q13.2

C/Superfamily: hepsin; trypsin homology

C/Keywords: hydrolase; liver; serine proteinase; transmembrane protein

F:23-45/Domain: transmembrane #status predicted <TN>

F:163-400/Domain: trypsin homology <TRY>

F:189-204,291-359,322-338,349-381/Disulfide bonds: #status predicted

F:203,257,353/Active site: His, Asp, Ser #status predicted

Query Match 31.6%; Score 505.5; DB 1; Length 417;

Best Local Similarity 41.1%; Pred. No. 1.5e-37;

Matches 101; Conservative 36; Mismatches 94; Indels 15; Gaps 4;

Qy 50 DKIVGGYALVDSPMQVSIQYDKQHVCGGSIILDPHWLTAACHFRKHTDVFN-WKYRAG 108

Db 161 DRIVGGRTSLGRNPQVSLRYDGAHLGSGSLSGDWLTAACHFRPERNVLSRWRFAG 220

Qy 109 SDKLGSPFSLAVAKIIIEFNPMYP-----KNDIALMKLQFPLTFSGTVRPICLPF 160

Db 221 AVAQASPHGLQLGVQAVVYHGGVLPFRDPNSENSNDIALVHLSSPLPLEYIQVCLPA 280

Qy 161 FDELTATPLNIGWGFTKQNGKMSDILLOASQVVDSTRCNADDAVOGEVTERKMC 220

Db 281 AGQALVDGKICTVTWGNT-QYTGQAGVLQEARVPIISNDVCNGADFYGNQIKPMFCA 339

Qy 221 GIPRGGVDTCCGSGGRLMYQ-----SDQHHVGVISWVGCGGSPSPGVVTVKVSALNW 275

Db 340 GYPEGGIDAQGGSGGFGFVCEDSISRTPRWRWLGIVSWGTCALAQKPGVTVKVSDFRE 399

Qy 276 IYNVWK 281

Db 400 IFQAIA 405

RESULT 2

JX0172

```

crosin (EC 3.4.21.10) precursor form 3 - mouse
;Species: Mus musculus (house mouse)
;Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 20-Jun-2000
;Accession: JX0172; JX0138
;Watanabe, K.; Baba, T.; Kashiwabara, S.; Okamoto, A.; Arai, Y.
;Biochem. 109, 828-833, 1991
;Title: Structure and organization of the mouse acrosin gene.
;Reference number: JX0172; MUID:92041732; PMID:1939002
;Accession: JX0172
;Molecule type: DNA
;Residues: 1-436 <WAT>
;Cross-references: NID:g238706; PIDN:AAB20293.1; PID:g238707
;Kashiwabara, S.; Baba, T.; Takada, M.; Watanabe, K.; Iano, Y.; Arai, Y.
;Biochem. 108, 785-791, 1990
;Title: Primary structure of mouse proacrosin deduced from the cDNA sequence and its ge
;Reference number: JX0138; MUID:91185535; PMID:2127931
;Accession: JX0138
;Molecule type: mRNA
;Residues: 4-436 <KAS>
;Cross-references: GS:D00754; NID:g220322; PIDN:BAA00651.1; PID:g220323
;Comment: Acrosin is an acrosomal protease that plays an important role in the initial
-binding activity.
;Genetics:
;Introns: 26/2; 95/2; 190/1; 238/3
;Superfamily: acrosin; trypsin homology
;Keywords: glycoprotein; hydrolase; serine proteinase; sperm
;1-19/Domain: signal sequence #status predicted <SIG>
;20-42/Product: acrosin light chain #status predicted <ALC>
;43-321/Product: acrosin heavy chain #status predicted <AHC>
;43-286/Domain: trypsin homology <TRY>
;22-211/Binding site: carbohydrate (Asn) (covalent) #status predicted
;25-155,29-162,74-90,178-247,210-226,237-267/Disulfide bonds: #status predicted
;89,143,241/Active site: His, Asp, Ser #status predicted
;235/Binding site: substrate (Asp) #status predicted

```

Query Match	31.2%	Score	499;	DB 2;	Length	436;			
Best Local Similarity	39.0%;	Pred.	No. 5.9e-37;						
Matches	112;	Conservative	43;	Mismatches	104;	Indels	28;	Gaps	9;
y	14	LLLVSNLLLCQGVSDYKDDDDVDAALAAAFDDDD-----KIVGGYALDVDSWPFOWSI	69						
b	4	MLPTVAVLVLAVSVA--KNITTCDCGPC-GLFRQNSQAGTIVSSQAQLGANFWVSIL	60						
y	70	OY-----DKOHVCGSGILDPHWLTAAACFRKHDTVDENWK-----VRAGSKLGSFP	116						
b	61	QITSHNSRRIHACGSLNSHWLTPAACFCDFNKKKYIDRWLVFGAQBIETGRNKPKEP	120						
y	117	SLA--VAKIIIBENPMYPKDNIALMKQFPLTTSGTVRPICLPFFDEETPATP--LW	172						
b	121	QQRYVQKIWIHEKYNVVTGEINDIALTKITPVPCGNFIGPCCLPHF-KAGSPQIHTCY	179						
y	173	IIGWGFTKQNGKMSDLILQASVOYLDSSTECNADDAYQGEVTEKMKWCAGIEPGGVDTCQG	232						
b	180	VTGWGYIKERAPPSPYLMEARVDLIDLDCNSTQWYNGRVISTNVVCAGYEGEKIDTCQG	239						
y	233	DSGGPLMYQSD---QNHVVGIWSVGCGGPGSTPPGYTTKVSAYLNW I	276						
b	240	DSGGPLMCRDNVDSPFVYVGGITTSWGVCARAKRPYGVIATMDYLDWI	286						

```

RESULT 3
37344      precursor form 1 - mouse
37344      crosin (EC 3.4.21.10)
;Species: Mus musculus (house mouse)
;Date: 19-Mar-1993 #sequence_revision 19-Mar-1993 #text_change 22-Jun-1999
;Accession: A37344      W.M.; Tsaousidou, S.; Adham, I.M.; Willison, K.; Engel, W.
;Klemm, U.; Maier, W.M.; Differentiation 42, 160-166, 1990
;Title: Mouse preproacrosin: cDNA sequence, primary structure and postmeiotic expression
;Reference number: A37344; MUID:90255839; PMID:2111255
;Accession: A37344
;Status: preliminary
;Molecule type: mRNA

```

A,Residues: 1-418 <KLE>
A,Cross-references: GB:X52466; NID:g49857; PIDN:CRA36704.1; PID:g49858
C:Superfamily: acrosin; trypsin homology
C:Keywords: glycoprotein; hydrolase; serine proteinase
F;39-283/Domain; trypsin homology <TRY>
F;18,208/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;21-151/Disulfide bonds: #status predicted
F;25-158/Disulfide bonds: #status predicted
F;70-86/Disulfide bonds: #status predicted
F;85,139,238/Active site: His, Asp, Ser #status predicted
F;175-244/Disulfide bonds: #status predicted
F;207-223/Disulfide bonds: #status predicted
F;234-264/Disulfide bonds: #status predicted

[illegible]

RESULT 4
KQWSPL
plasma kallikrein (EC 3.4.21.34) precursor - mouse
C/Species: Mus musculus (house mouse)
C/Date: 30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change 18-Jun-1999
C/Accession: A36557
R/Seidsh, N.G.; Sawyer, N.; Hamelin, J.; Mion, P.; Beaubien, G.; Brachpapa, L.; Rochement
DNA Cell Biol. 9, 737-748, 1990
A/Title: Mouse plasma kallikrein: cDNA structure, enzyme characterization, and comparison
A/Reference number: A36557; MUID:91090844; PMID:2264928
A/Accession: A36557
A/Molecule type: mRNA
A/Residues: 1-638 <SEI>
A/Cross-references: GB:M58588; NID:G200358; PIDN:AAA63393.1; PID:G200359
A/Note: part of this sequence, including the amino ends of both the heavy and light chain
C/Comment: This protein, synthesized in the liver, circulates as a noncovalent complex w
C/Comment: The zymogen is activated by factor XIIa, which cleaves the molecule into a lig
are linked by one or more disulfide bonds.
C/Superfamily: coagulation factor XI; trypsin homology
C/Keywords: blood coagulation; duplication; fibrinolysis; glycoprotein; hydrolase; infla
F/1-190/Domain: signal sequence #status predicted <SIG>
F/20-390/Product: plasma kallikrein heavy chain #status experimental <HCH>
F/20-109/Domain: apple repeat <AP1>
F/110-199/Domain: apple repeat <AP2>
F/200-289/Domain: apple repeat <AP3>
F/291-380/Domain: apple repeat <AP4>
F/391-638/Product: plasma kallikrein light chain #status experimental <LCH>
F/391-621/Domain: trypsin homology <TRY>
F/21-104,47-77,51-57,111-194,137-166,141-147,201-284,227-256,231-237,292-375,318-347,322-
F/127-215,308,396,494/Binding site: carbohydrate (Asn) (covalent) #status predicted
F/434,483,578/Active site: His, Asp, Ser #status predicted
Query Match 30.8%; Score 492; DB 1; Length 638;
Best Local Similarity 43.5%; Pred. No. 4e-36;

QY 70 QY-----DKQHVGCGSITLDPHHWLVTAAHCFPKKHTDVFNNK-----VRAGSDKLGSFP 116
Db :
61 QIFTSNRRRYRHACGSGILLNSHHWLVLTAAHCFDNKKVVDWRVLVFGAHEIYEYGNKPVKPE 120
QY 117 SLA--VAKIIIEENPMYPKNDIALMKLOPLTFSGTVRPICLPFDDEBLTPATP--LW 172
Db :
121 QQERYVQKVIHEKYNAVTEGNDIALLVPTTVPVTCGDFVGPCLPHF-KSGPPRIPTHCY 179
QY 173 IIGWGFTKQNGKGMSDILLOASVOVIDSTRCNADDAYOQEVEKMKCAGIPGGVDTCQG 232
Db :
180 VTGWGYIKDNAEPSPVLMREARVLIDLDLNCSTWFYNGRVSTNVCAGYPEGKIDTCQG 239
QY 233 DSGGPLMYQSDQ---WHVGVIVSWGYGCGGSTPGVYTKVSAYLNWI 276
Db :
240 DSGGPLMCRDTRRQPFVIVGITSGVGCARAKRPGVYTATWDYLDWI 286

RESULT 6

KQRTPL
plasma kallikrein (EC 3.4.21.34) precursor - rat
N;Alternate names: Fletcher factor; kininogenin; serum kallikrein
C;Species: Rattus norvegicus (Norway rat)
C;Date: 30-Sep-1992 #sequence, revision 30-Sep-1992 #text_change 18-Jun-1999
C;Accession: A39180; A33320; S66851; I53041; S06852
R;Beaubien, G.; Rosinski-Chupin, I.; Mattei, M.G.; Mbikay, M.; Chretien, M.; Seidah, N.G.
Biochemistry 30, 1628-1635, 1991
A;Title: Gene structure and chromosomal localization of plasma kallikrein.
A;Reference number: A39180; MUID:91129236; PMID:1993180
A;Accession: A39180
A;Molecule type: DNA
A;Residues: 1-638 <BEA>
A;Cross-references: GB:J05315
A;Note: The authors translated the codon GAG for residue 81 as Glu
R;Seidah, N.G.; Ladenheim, R.; Mbikay, M.; Hamelin, J.; Lutfalla, G.; Rougeon, F.; Lazure,
DNA 8, 563-574, 1989
A;Title: The cDNA structure of rat plasma kallikrein.
A;Reference number: A33320; MUID:90091743; PMID:259771
A;Accession: A33320
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-638 <SEI>
A;Cross-references: GB:M30282; NID:g205010; PIDN:AAA41463.1; PID:g205011
A;Note: part of this sequence, including the amino ends of both the heavy and light chain
R;Paquin, J.; Benjannet, S.; Sawyer, N.; Lazure, C.; Chretien, M.; Seidah, N.G.
Biochim. Biophys. Acta 999, 103-110, 1989
A;Title: Rat plasma kallikrein: purification, NH(2)-terminal sequencing and development
A;Reference number: S06851; MUID:90089457; PMID:2597701
A;Accession: S06851
A;Molecule type: protein
A;Residues: 20-45;351-413 <PAQ>
R;Seidah, N.G.; Ladenheim, R.; Mbikay, M.; Hamelin, J.; Lutfalla, G.; Rougeon, R.; Lazure,
DNA Cell Biol. 8, 563-574, 1989
A;Title: The cDNA structure of rat plasma kallikrein.
A;Reference number: I53041
A;Accession: I53041
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-638 <RES>
A;Cross-references: GB:M58590; NID:g206721; PIDN:AAA42069.1; PID:g206722
C;Comment: This protein, synthesized in the liver, circulates as a noncovalent complex with
C;Comment: The zymogen is activated by factor XIa, which cleaves the molecule into a lig
are linked by one or more disulfide bonds.
C;Genetics:
A;Gene: PK
C;Superfamily: coagulation factor XI; trypsin homology
C;Keywords: blood coagulation; duplication; fibrinolysis; glycoprotein; hydrolase; inflan
F;1-19/Domain: signal sequence #status predicted <SIG>
F;20-390/Product: plasma kallikrein heavy chain #status experimental <MAT1>
F;20-109/Domain: apple repeat <AP1>
F;110-199/Domain: apple repeat <AP2>
F;200-289/Domain: apple repeat <AP3>
F;291-380/Domain: apple repeat <AP4>
F;391-638/Product: plasma kallikrein light chain #status experimental <MAT2>

```
;391-621/Domain: trypsin homology <TRY>
;21-104,47-77,51-57,111-194,137-166,141-147,201-284,227-256,231-237,292-375,318-347,322
;127,215,308,453,459,494/Binding site: carbohydrate (Asn) (covalent) #status predicted
;396/Binding site: carbohydrate (Asn) (covalent) #status experimental
;434,483,578/Active site: His, Asp, Ser #status predicted

Query Match      30.6%; Score 489; DB 1; Length 638;
Best Local Similarity 42.2%; Pred. No. 7.4e-36;
Matches 100; Conservative 45; Mismatches 76; Indels 16; Gaps 8;

Y 51 KIVGGVALDVSDPWPQVSIQ---YDKQHVCGSILDPHWLTAHCF--KHTDVFNVKV 105
b 390 RIVGGTSSNGEWPQVSLQVLTAKHCGSLIGHQWVLTAAHCFDGLPQDV--WRI 447
Y 106 RAG----SDKLGSPFLSAVAKIIIEFNPMYKNDIALMKLOFPLTFSGTVRPCLPFF 161
b 448 YGILNLSEITNKTFFSSIKELIIHQYKMSGSDYDIALIKLQTLNLTFFOKPCLPSK 507
Y 162 DEELTPATPLNIIGWGTQKNGKMSDILLQASVQVIDSTRCNADDAVQGEV-ETEKMCA 220
b 508 ADTNTIYTNVCWVTGWGTYKER-GETQNLQKATIPLVNEEC--QKKYRDYVITKQICA 564
Y 221 GIPEGVDTCQDGGGLMYQ-SDQHVHVGVISWYCGCGPSTPGYTKVSAYLNIW 276
b 565 GYKEGGIDACKDGGPLVCKHSGRWLGVITSGEGCARKEQPGYTKVAEYDNIW 621

RESULT 7
33777
;Species: Rattus norvegicus (Norway rat)
;Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 18-Jun-1999
;Accession: S33777; S32013
;Farley, D.; Raymond, F.; Nick, H.
;iochim. Biophys. Acta 1173, 350-352, 1993
;Title: Cloning and sequence analysis of rat hepsin, a cell surface serine proteinase.
;Reference number: S33777; PMID:93305733; PMID:8318546
;Accession: S33777
;Status: preliminary
;Molecule type: mRNA
;Residues: 1-416 <FAR>
;Cross-references: EMBL:X70900; NID:g57928; PIDN:CAAS0256.1; PID:g57929
;Superfamily: hepsin; trypsin homology
;Keywords: hydrolase; liver; serine proteinase; transmembrane protein
;22-44/Domain: transmembrane #status predicted <TM>
;162-339/Domain: trypsin homology <TRY>
;187-203,290-358,321-337,348-380/Disulfide bonds: #status predicted
;202,256,352/Active site: His, Asp, Ser #status predicted

Query Match      30.0%; Score 479.5; DB 1; Length 416;
Best Local Similarity 39.8%; Pred. No. 3.2e-35;
Matches 98; Conservative 38; Mismatches 95; Indels 15; Gaps 4;

Y 50 DKIVGGVALDVSDPWPQVSIQYDKQHVCGSILDPHWLTAHCFKRTDYFN-WKVRAG 108
b 160 DRIVGGQSSLGSRPWPQVSLRYDGHLCGSLSGDWLTAHCFEPNRLSRVFEAG 219
Y 109 SDKLGSPFLSAVAKIIIEFNPMYP-----KNDIALMKLOFPLTFSGTVRPCLPFF 160
b 220 AVARTSPHAVQLGQVAVIHGGYLPFRDPTDENSNDIALVELSSSLPTIYIQLVLA 279
Y 161 FDELTPTATPLNIIGWGTQKNGKMSDILLQASVQVIDSTRCNADDAVQGEVETEKMCA 220
b 280 AQQALVDGKVTCTVGNT-QFYGQAVQLQEARVPIISNEVQNSPDFYGNQIKPKNFCA 338
Y 221 GIPEGVDTCQDGGGLMYQ-----SDQHVHVGVISWYCGCGPSTPGYTKVSAYLNIW 275
b 339 GYPEGGIDACQDGGHFPVCEDRISGTSRWRLCGIVSGWTGCLARKEPGYTKVIDFREW 398
Y 276 IYNVK 281
b 399 IFQAIK 404
```

RESULT 8

KOHUP

Plasma kallikrein (EC 3.4.21.34) precursor - human

N:Alternate names: kininogenin; plasma prekallikrein

C:Species: Homo sapiens (man)

C>Date: 13-Aug-1986 #sequence_revision 13-Aug-1986 #text_change 18-Jun-1999

C:Accession: A00921; A37939

C:Chung, D.W.; Fujikawa, K.; McMullen, B.A.; Davie, E.W.

Biochemistry 25, 2410-2417, 1986

A:Title: Human plasma prekallikrein, a zymogen to a serine protease that contains four tr

A:Reference number: A00921; PMID:86243359; PMID:3521732

A:Accession: A00921

A:Molecule type: mRNA

A:Residues: 1-638 <CHU>

A:Cross-references: GB:M13143; NID:g190262; PIDN:AAA60153.1; PID:g190263

R:McMullen, B.A.; Fujikawa, K.; Davie, E.W.

Biochemistry 30, 2050-2056, 1991

A:Title: Location of the disulfide bonds in human plasma prekallikrein: the presence of f

A:Reference number: A37939; PMID:91152016; PMID:1998666

A:Accession: A37939

A:Molecule type: protein

A:Residues: 20-27,40-46, 'X', 48, 'H', 50, 'X', 52-70, 'H', 75-76, 'X', 78-80, 103-113, 131-140, 141-1

;260-283, 'X', 285;287-291, 'X', 293-295;314-317, 'X', 319-320;321-324, 'X', 329-333;334-339, 'X',

525;538-551;562, 'X', 564-567;573, 'X', 575-576;578-583, 'X', 585;592-604 <MCM>

C:Comment: This protein, synthesized in the liver, circulates as a noncovalent complex w

C:Comment: The zymogen is activated by factor XIIa, which cleaves the molecule into a lig

are linked by one or more disulfide bonds.

C:Comment: The enzyme cleaves Lys-Arg and Arg-Ser bonds. It activates, in a reciprocal r

inogen and may also play a role in the renin-angiotensin system by converting protein in

C:Genetics:

A:Gene: GDB:KLIK3

A:Cross-references: GDB:127575; OMIM:229000

A:Map position: 4q35-q35

C:Superfamily: coagulation factor XI; trypsin homology

C:Keywords: blood coagulation; duplication; fibrinolysis; glycoprotein; hydrolase; inflan

F:1-19/Domain: signal sequence #status predicted <SIG>

F:20-638/Product: plasma kallikrein #status predicted <MAT>

F:20-390/Domain: plasma kallikrein heavy chain #status predicted <HCH>

F:20-109/Domain: apple repeat <AP1>

F:110-199/Domain: apple repeat <AP2>

F:200-289/Domain: apple repeat <AP3>

F:291-380/Domain: apple repeat <AP4>

F:391-638/Domain: plasma kallikrein light chain #status predicted <LCH>

F:21-104,47-77,51-57,111-194,137-166,141-147,201-284,227-256,231-237,292-375,322-328,383-

F:127,308,396,453,494/Binding site: carbohydrate (Asn) (covalent) #status experimental

F:318-347,340-345/Disulfide bonds: #status predicted

F:390-391/Cleavage site: Arg-Ile (coagulation factor XIIa) #status predicted

F:434,483,578/Active site: His, Asp, Ser #status predicted

Query Match 29.9%; Score 479; DB 1; Length 638;

Best Local Similarity 41.8%; Pred. No. 5.8e-35;

Matches 99; Conservative 46; Mismatches 76; Indels 16; Gaps 8;

QY 51 KIVGGVALDVSDPWPQVSIQYD---KQHVCGSILDPHWLTAHCF--RKHTDVFNVKV 105

Db 390 RIVGGTSSNGEWPQVSLQVLTAKHCGSLIGHQWVLTAAHCFDGLPQDV--WRI 447

QY 106 RAG----SDKLGSPFLSAVAKIIIEFNPMYKNDIALMKLOFPLTFSGTVRPCLPFF 161

Db 448 YGILNLSDITKDTFFSQIKELIIHQYKVSSEGNHDIKILQAPLNYTEFQKPLPSK 507

QY 162 DEELTPATPLNIIGWGTQKNGKMSDILLQASVQVIDSTRCNADDAVQGEV-ETEKMCA 220

Db 508 GDTSTIYTNVCWVTGWGFSKEK-GEIQNLQKVNIPLVNTEEC--QKKYQDYKITQRMVCA 564

QY 221 GIPEGVDTCQDGGGLMYQ-SDQHVHVGVISWYCGCGPSTPGYTKVSAYLNIW 276

Db 565 GYKEGGIDACKDGGPLVCKHNGMRLVIGITSGEGCARKEQPGYTKVAEYDNIW 621

RESULT 9

A;Cross-references: EMBL:X54017
R;Adham, I.M.; Klemm, U.; Maier, W.M.; Engel, W.
Hum. Genet. 84, 125-128, 1990
A;Title: Molecular cloning of human preproacrosin cDNA.
A;Reference number: A61022; MUID:90128986; PMID:2298447
A;Accession: A61022
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
R;Baba, T.; Watanabe, K.; Kashiwabara, S.I.; Arai, Y.
FEBS Lett. 244, 296-300, 1989
A;Title: Primary structure of human proacrosin deduced from its cDNA sequence.
A;Reference number: S03330; MUID:89153568; PMID:2493394
A;Accession: S03330
A;Molecule type: mRNA
A;Residues: 1-63, 'T', 65-119, 'V', 121-165, 'L', 167-267, 'R', 269-344, 'R', 346-421 <BAB>
A;Cross-references: EMBL:Y00970; NID:928225; PIDN:CAA60784.1; PID:928326
C;Genetics:
A;Gene: GDB:ACR
A;Cross-references: GDB:119645; OMIM:102480
A;Map position: 22q13-22qter
A;Introns: 26/2; 94/2; 189/2; 237/3
C;Superfamily: acrosin; trypsin homology
C;Keywords: glycoprotein; hydrolase; serine proteinase; sperm
F;1-19/Domain: signal sequence #status predicted <SIG>
F;20-421/Product: acrosin #status predicted <MAR>
F;20-42/Product: acrosin light chain #status predicted <LCH>
F;43-421/Product: acrosin heavy chain #status predicted <HCH>
F;43-285/Domain: trypsin homology <TRY>
F;302-379/Region: proline-rich
F;22,210/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;25-154/Disulfide bonds: #status predicted
F;29-162/Disulfide bonds: #status predicted
F;73-89/Disulfide bonds: #status predicted
F;88,142,240/Active site: His, Asp, Ser #status predicted
F;177-246/Disulfide bonds: #status predicted
F;209-225/Disulfide bonds: #status predicted
F;236-266/Disulfide bonds: #status predicted

Query Match 29.88; Score 476; DB 1; Length 421;
Best Local Similarity 41.08; Pred. No. 6.6e-35;
Matches 100; Conservative 37; Mismatches 89; Indels 18; Gaps 5;

QY 51 KIVGGYALDVDSFPWQVSIQY-----DKQHVCGGSIIDPHWLVTAACHCRKHTDVFNNKV 105
Db 42 RIVGKRAAQHGAFWVMVSLQIFRNSHRYHTCGSLLNSRWLVTAACHFVGKNNVHDWEL 101
QY 106 RAGSDKL---GSFPSLA-----VAKIIIEFNPMYPKNDIALMKLQFPLTFSGTVRPI 156
Db 102 VFGAKETTYGNKFPVKAPLQERVVEKIIIEKNSATEGNDIALVETTPISCGRFIFGP 161
QY 157 CLPFFDEELTPAT-PLWIIIGWTKNGKMSDILLQASQVIVDTRCNADDAQGEYTE 215
Db 162 CLPFFXAGLPRGSCQWAGVIEKAPRPSILLMEARVLDLIDLCNLTQWYNGRVQP 221
QY 216 KMCAGIPEGGVDTCCGDSGGLMYQSDQ---VHWVGIVSWGVCGGPSTPGVYTKVSAY 272
Db 222 TNVCAGYFVGKIDTCQDSGGLMCKDKSKESAYVVGILTSWVGICALAKRGIYATWYF 281
QY 273 LNWI 276
Db 282 LNWI 285

RESULT 11
A34170
N;Alternate names: 53K fucose-binding protein
C;Species: Sus scrofa domestica (domestic pig)
C;Date: 10-Sep-1999 #sequence revision 10-Sep-1999 #text change 10-Sep-1999
C;Accession: A34170; S08994; S02428; S04940; S16657; S02780; S10695; S12968
R;Baba, T.; Kashiwabara, S.; Watanabe, K.; Itoh, H.; Michikawa, Y.; Kimura, K.;
J. Biol. Chem. 264, 11920-11927, 1989.

Query Match	29.6%;	Score 473.5;	DB 1;	Length 415;
Best Local Similarity	41.5%;	Pred. No. 1.1e-34;		

b 51 SLISEQWVVSAGHCYKSRIOVLGEHNIEVLGNEQF-----INAKII-----RHPKY 99
y 135 -----DNDIALMKLOPPLTFSGTVRPICLPFFFEELTPA--TPLWIIIGWFTKQGGKMS 187
b 100 NSRTLDNDILLIKLSSPAVINSRVSAISLP-----TAPPAAGTESLISGWENTLSSGADYP 155
y 188 DILLQASVQVIDSTRCNADDAIYQGEVTERMCAGIPEGGVDTCCGDSGGPLMYQSDQWHV 247
b 156 DELQCLDAFVLSQAECEA--SYFGKITNNMFCVGFLEGGKDSQQDSGGPVVVSNGE--L 210
y 248 VGIVSWGVCGGGSTPGVYTKYSAYLNWI 276
b 211 QGIVSWGICGAQXNRPGVYTKVINYVDWI 239

earch completed: November 23, 2003, 07:52:54
ob time : 20.2792 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

M protein - protein search, using sw model

un on: November 23, 2003, 07:39:05 ; Search time 11.2462 Seconds
(without alignments)
1221.015 Million cell updates/sec

itle: US-09-607-745-9

erfect score: 1600

equences: 1 MDSKSSQKSLLLLLLVSN.....LNWIYVWKAELSRHHHHH 292

coring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

earched: 127863 seqs, 47026705 residues

otal number of hits satisfying chosen parameters: 127863

inimum DB seq length: 0

aximum DB seq length: 2000000000

ost-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

atabase : SwissProt_41:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

result No.	Score	Query Match %	Length	ID	Description
1	1266	79.1	437	1 TMS4_HUMAN	Q9rs4 homo sapien
2	526.5	32.9	418	1 HATT_HUMAN	O60235 homo sapien
3	523	32.7	453	1 TMS3_MOUSE	Q8kl0 mus musculus
4	517	32.3	455	1 TMS5_MOUSE	Q9er04 mus musculus
5	509.5	31.8	454	1 TMS3_HUMAN	P57727 homo sapien
6	506	31.6	492	1 TMS2_HUMAN	O15393 homo sapien
7	505.5	31.5	417	1 HEPS_HUMAN	P05981 homo sapien
8	499	31.2	436	1 ACRO_MOUSE	P23578 mus musculus
9	492	30.8	638	1 KAL_MOUSE	P26262 mus musculus
10	491.5	30.7	435	1 HEPS_MOUSE	O35453 mus musculus
11	491	30.7	437	1 ACRO_RAT	P29293 rattus norv
12	489	30.6	638	1 KAL_RAT	P14272 rattus norv
13	487	30.4	457	1 TMS5_HUMAN	Q9h3s3 homo sapien
14	483	30.2	324	1 TEST_MOUSE	Q9lhj7 mus musculus
15	480.5	30.0	321	1 TRYG_HUMAN	Q9nr12 homo sapien
16	479.5	30.0	416	1 HEPS_RAT	Q05511 rattus norv
17	479	29.9	638	1 KAL_HUMAN	P03952 homo sapien
18	476	29.8	421	1 ACRO_HUMAN	P10323 homo sapien
19	476	29.8	490	1 TMS2_MOUSE	Q9j1q8 mus musculus
20	475.5	29.7	311	1 ACRO_MOUSE	Q9l07 mus musculus
21	473.5	29.6	415	1 TRYG_MOUSE	P08001 mus musculus
22	470.5	29.4	246	1 TRY2_MOUSE	P07146 mus musculus
23	469	29.3	277	1 TRY2_ANOGA	P07478 homo sapien
24	464.5	29.0	247	1 TRY1_RAT	P00762 rattus norv
25	463.5	29.0	246	1 TRY7_ANOGA	P35041 anopheles g
26	463.5	29.0	267	1 ACRO_RABIT	P48038 cryptolegus
27	461.5	29.0	431	1 ACRO_RABIT	Q9ul52 homo sapien
28	459.5	28.8	422	1 DES1_HUMAN	P06872 canis famil
29	459.5	28.7	247	1 TRY2_CANIFA	P97435 mus musculus
30	458	28.6	1069	1 ENTK_MOUSE	P9es87 rattus norv
31	457	28.6	342	1 PSS8_RAT	P35037 anopheles g
32	456.5	28.5	275	1 TRY3_ANOGA	P00763 rattus norv
33	451.5	28.2	246	1 TRY2_RAT	

RESULT 1
TMS4_HUMAN 314 28.2 451.5 28.1 247 1 TEST_HUMAN
ID TMS4_HUMAN STANDARD; PRT; 437 AA.
AC Q9NRS4; Q9NZA5;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Transmembrane protease, serine 4 (EC 3.4.21.-) (Membrane-type serine
protease 2) (MT-SP2).
GN TMPSRS4 OR TMPSRS3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
SEQUENCE FROM N.A.
RP TISSUE=Pancreatic carcinoma;
RC MEDLINE=20283276; PubMed=10825129;
RA Wallrapp C., Haehnel S., Mueller-Pillasch F., Burghardt B.,
RA Iwamura T., Ruthenbuenger W., Lerch M.M., Adler G., Gress T.M.;
RT "A novel transmembrane serine protease (TMPSRS3) overexpressed in
pancreatic cancer.",
RL Cancer Res. 60:2602-2606(2000).
[2]
SEQUENCE FROM N.A.
RP Smeekens S.S., Lorrimer D.D., Wang E., Hou J., Linnevers C.;
RT "MT-SP2, a novel type II membrane serine protease expressed in
trachea, colon, and small intestine: identification, cloning, and
chromosomal localization.",
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
[3]
SEQUENCE FROM N.A.
RP TISSUE=Pancreas;
RC MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Casavant T.L., Scheetz T.E.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Rubin G.M., Hong L.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Murley D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences.",
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC -!- FUNCTION: MAY PLAY A ROLE IN TUMOR PROGRESSION, METASTASIS
FORMATION AND TUMOR INVASION.

Q9y6m0 homo sapien
P07477 homo sapien
Q9esd1 mus musculus
P20918 mus musculus
P56677 mus musculus
P35035 anopheles g
P35030 homo sapien
Q9y5y6 homo sapien
P00760 bos taurus
Q29463 bos taurus
P06867 sus scrofa
P00761 sus scrofa

ALIGNMENTS

C -1- SUBCELLULAR LOCATION: Type II membrane protein (potential).
C -1- TISSUE SPECIFICITY: HIGH LEVELS IN PANCREATIC, GASTRIC, COLORECTAL
C AND AMPULLARY CANCER. VERY WEAK EXPRESSION IN NORMAL
C GASTROINTESTINAL AND UROGENITAL TRACT.
C -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
C -1- SIMILARITY: Contains 1 SRCR domain.
C -1- SIMILARITY: Contains 1 LDL-receptor class A domain.
C
C This SWISS-PROT entry is copyright. It is produced through a collaboration
C between the Swiss Institute of Bioinformatics and the EMBL outstation -
C the European Bioinformatics Institute. There are no restrictions on its
C use by non-profit institutions as long as its content is in no way
C modified and this statement is not removed. Usage by and for commercial
C entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
C or send an email to license@isb-sib.ch).
C
C R EMBL: AF179224; AAF74526.1; -.
C R EMBL: AF216312; AAF31436.1; -.
C R EMBL: BC011703; AAH11703.1; -.
C R HSP: P00763; IDPO.
C R MROPS; S01.034; -.
C R Genew; HGNC:11878; TMPSR54.
C R MIM; 606565; -.
C R GO; GO:0016021; C: integral to membrane; NAS.
C R GO; GO:0006508; P: proteolysis and peptidolysis; NAS.
C R InterPro; IPR001314; Chymotrypsin.
C R InterPro; IPR002172; LDL receptor A.
C R InterPro; IPR001254; Ser_protease_Try.
C R InterPro; IPR001190; Srcr_receptor.
C R Pfam; PF00057; ldl_recept_a; 1.
C R Pfam; PF00089; trypsin; 1.
C R PRINTS; PR00722; CHYMOTRYPSIN.
C R SMART; SM00192; LDLA; 1.
C R SMART; SM00202; SR; 1.
C R SMART; SM00202; Tryp_Spc; 1.
C R PROSITE; PS01209; LDLRA_1; FALSE NEG.
C R PROSITE; PS00068; LDLRA_2; FALSE NEG.
C R PROSITE; PS00420; SRCR_1; FALSE NEG.
C R PROSITE; PS00287; SRCR_2; 1.
C R PROSITE; PS00240; TRYPSIN_DOM; 1.
C R PROSITE; PS00134; TRYPSIN_HIS; 1.
C R PROSITE; PS00135; TRYPSIN_SER; 1.
C R Hydrolase; Serine protease; Transmembrane; Signal-anchor.
C W DOMAIN 1 38 CYTOPLASMIC (POTENTIAL).
C T SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
C T TRANSMEM 39 59 (POTENTIAL).
C T DOMAIN 60 437 EXTRACELLULAR (POTENTIAL).
C T LDL-RECEPTOR CLASS A.
C T SRCR.
C T SERINE PROTEASE.
C T CHARGE RELAY SYSTEM (BY SIMILARITY).
C T CHARGE RELAY SYSTEM (BY SIMILARITY).
C T CHARGE RELAY SYSTEM (BY SIMILARITY).
C T CLEAVAGE (POTENTIAL).
C T BY SIMILARITY.
C T BY SIMILARITY.
C T BY SIMILARITY.
C T BY SIMILARITY.
C T BY SIMILARITY.
C T BY SIMILARITY.
C T N-LINKED (GLCNAC...) (POTENTIAL).
C T N-LINKED (GLCNAC...) (POTENTIAL).
C T MLOPDSQDPLNSIDVPLKRPIMETFEK -> MSNPCA
C T NPVSFWPSPS (IN REF. 2).
C T CONFLICT 1 31
C T SEQUENCE 437 AA; 48204 MW; 351B2FD4A8657B12 CRC64;
C
C Query Match 79.1%; Score 1266; DB 1; Length 437;
C Best Local Similarity 97.4%; Pred. No. 5.5e-111;
C Matches 228; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
C Y 51 KINGGYALDVDSWEPWQVSIQYDKQVCGSILDPHWLTAACHFRKHTDVFNKVRAGSD 110

Db 204 RVVGEASVDSWEPWQVSIQYDKQVCGSILDPHWLTAACHFRKHTDVFNKVRAGSD 263
Qy 111 KLGSPFLSLAVAKIIIIIFENPMYKPDNDIALMKLQFLFTSGTVRPICLPFFDELTATP 170
Db 264 KLGSPFLSLAVAKIIIIIFENPMYKPDNDIALMKLQFLFTSGTVRPICLPFFDELTATP 323
Qy 171 LWIIGWFTQNGKSGDILLOASVQVIDSTRCNADDAVQGEVTERKMCAGIEGGVDTC 230
Db 324 LWIIGWFTQNGKSGDILLOASVQVIDSTRCNADDAVQGEVTERKMCAGIEGGVDTC 383
Qy 231 QDSDGGPLMQSDQWVGVVSWGYGGGSGTSGVTKVSAYLNWVNWKAEL 284
Db 384 QDSDGGPLMQSDQWVGVVSWGYGGGSGTSGVTKVSAYLNWVNWKAEL 437

RESULT 2
HATT HUMAN
ID HATT HUMAN STANDARD; PRT; 418 AA.
AC O60235;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE 28-FEB-2003 (Rel. 41, Last annotation update)
DE Airway trypsin-like protease precursor (EC 3.4.21.-).
GN HAT.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98234382; PubMed=9565616;
RA Yamaoka K., Masuda K.-I., Ogawa H., Takagi K., Umemoto N., Yasuoka S.;
RT "Cloning and characterization of the cDNA for human airway trypsin-
RT like protease.",
RL J. Biol. Chem. 273:11895-11901(1998).
RN [2]
RP SEQUENCE OF 187-206, AND CHARACTERIZATION.
RX MEDLINE=97224034; PubMed=9070615;
RA Yasuoka S., Ohnishi T., Kawano S., Tsuchihashi S., Ogawara M.,
RA Masuda K.-I., Yamaoka K., Takahashi M., Sano T.;
RT "Purification, characterization, and localization of a novel
RT trypsin-like protease found in the human airway.",
RL Am. J. Respir. Cell Mol. Biol. 16:300-308(1997).
CC -1- FUNCTION: May play some biological role in the host defense system
CC on the mucous membrane independently of or in cooperation with
CC other substances in airway mucous or bronchial secretions.
CC -1- CATALYTIC ACTIVITY: Preferentially cleaves the C-terminal side of
CC arginine residues at the P1 position of certain peptides, cleaving
CC Boc-Phe-Ser-Arg-4-methylcoumaryl-7-amide most efficiently and
CC having an optimum pH of 8.6 with this substrate.
CC -1- ENZYME REGULATION: Strongly inhibited by diisopropyl
CC fluorophosphate, leupeptin, antipain, aprotinin, and soybean
CC trypsin inhibitor, but hardly inhibited by secretory leukocyte
CC protease inhibitor at 10 microm.
CC -1- SUBUNIT: Monomer.
CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. ACTIVATED BY
CC CLEAVAGE AND SECRETED.
CC -1- TISSUE SPECIFICITY: Located in the cells of the submucosal serous
CC glands of the bronchi and trachea.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
CC -1- SIMILARITY: Contains 1 SEA domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AB002134; BAA28691.1; -.
CC HSP: P00750; 1RTF.

390 LTGVDVSCQDGGGLVCQBRLWLKLVGATSPGICAEVNKPGVYTRVTSLFDWIHQE 444

Dd 282 AEL 284
Qy .!
Db 450 RDL 452

RESULT 6

TMS2 HUMAN STANDARD; PRT; 492 AA.

ID TMS2_HUMAN Q9BX11;
AC OL5393; Q9BX11;
DT 15-JUL-1998 (Rel. 36, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Transmembrane protease, serine 2 precursor (EC 3.4.21.-).
GN TMPRSS2 OR PRSS10.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9748144; PubMed=9325052;
RA Paoloni-Giacobino A., Chen H., Feitsch M.C., Rossier C.,
RA Antonarakis S.E.;
RT "Cloning of the TMPRSS2 gene, which encodes a novel serine protease
with transmembrane, LDLRA, and SRCR domains and maps to 21q22.3";
RL Genomics 44:309-320(1997).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=21309069; PubMed=11414763;
RA Teng D.H., Chen Y., Lian L., Ha P.C., Tavtigian S.V., Wong A.K.;
RT "Mutation analyses of 268 candidate genes in human tumor cell lines.";
RL Genomics 74:352-364(2001).
RN [3]
RP SEQUENCE FROM N.A., AND MUTAGENESIS.
RX MEDLINE=21139112; PubMed=11245484;
RA Afar D.E.H., Vivanco I., Hubert R.S., Kuo J., Chen E., Saffran D.C.,
RA Raitano A.B., Jakubovits A.;
RT "Catalytic cleavage of the androgen-regulated TMPRSS2 protease results
in its secretion by prostate and prostate cancer epithelia.";
RL Cancer Res. 61:1686-1692(2001).
RN [4]
RP TISSUE SPECIFICITY.
RX MEDLINE=21104370; PubMed=11169536;
RA Vaarala M.H., Porvari K.S., Kellokumpu S., Kyllonen A.P., Vibko P.T.;
RT "Expression of transmembrane serine protease TMPRSS2 in mouse and
human tissues.";
RL J. Pathol. 193:134-140(2001).
CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. ACTIVATED BY
CLEAVAGE AND SECRETED.
CC -1- TISSUE SPECIFICITY: EXPRESSED STRONGLY IN SMALL INTESTINE. ALSO
EXPRESSED IN PROSTATE, COLON, STOMACH AND SALIVARY GLAND.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
CC -1- SIMILARITY: Contains 1 SRCR domain.
CC -1- SIMILARITY: Contains 1 LDL-receptor class A domain.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).

EMBL; U75329; AAC51784.1; --
EMBL; AF123453; AAD37117.1; --
EMBL; AF270487; AAK29280.1; --
HSSP; P00763; 1DPO.
DR MEROPS; S01.247; --
Genev; HGNC:11876; TMPRSS2.
DR MIM; 602060; --
GO; GO:0005887; C:integral to plasma membrane; TAS.
DR DR

```

R InterPro; IPR001314; Chymotrypsin.
R InterPro; IPR002172; LDL_receptor_A.
R InterPro; IPR001254; Ser_protease_Try.
R InterPro; IPR001190; Sscr_receptor.
R Pfam; PF00089; trypsin; 1.
R PRINTS; PR00722; CHYMOTRYPSIN.
R SMART; SM00192; LDLa; 1.
R SMART; SM00202; SR; 1.
R SMART; SM00202; Tryp_Spc; 1.
R PROSITE; PS01209; LDLRA_1; 1.
R PROSITE; PS00068; LDLRA_2; 1.
R PROSITE; PS00420; SRCR_1; FALSE_NEG.
R PROSITE; PS00287; SRCR_2; 1.
R PROSITE; PS00240; TRYPSIN_DOM; 1.
R PROSITE; PS00134; TRYPSIN_HIS; 1.
R PROSITE; PS00135; TRYPSIN_SER; 1.
W Hydrolase; Serine protease; Transmembrane; Signal-anchor; Zymogen;
W Polymorphism.
T CHAIN 1 255 TRANSMEMBRANE PROTEASE, SERINE 2, NON-
T CATALYTIC CHAIN.
T CHAIN 256 492 TRANSMEMBRANE PROTEASE, SERINE 2,
T CATALYTIC CHAIN.
T DOMAIN 1 84 CYTOPLASMIC (POTENTIAL).
T TRANSMEM 85 105 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
(POTENTIAL).
T DOMAIN 106 492 EXTRACELLULAR (POTENTIAL).
T DOMAIN 112 149 LDL-RECEPTOR CLASS A.
T DOMAIN 150 242 SRCR.
T DOMAIN 256 492 SERINE PROTEASE.
T ACT_SITE 296 296 CHARGE RELAY SYSTEM (BY SIMILARITY).
T ACT_SITE 345 345 CHARGE RELAY SYSTEM (BY SIMILARITY).
T ACT_SITE 441 441 CHARGE RELAY SYSTEM.
T SITE 255 256 CLEAVAGE (POTENTIAL).
T DISULFID 113 126 BY SIMILARITY.
T DISULFID 120 139 BY SIMILARITY.
T DISULFID 133 148 BY SIMILARITY.
T DISULFID 172 231 BY SIMILARITY.
T DISULFID 185 241 BY SIMILARITY.
T DISULFID 244 365 INTERCHAIN (BY SIMILARITY).
T DISULFID 281 297 BY SIMILARITY.
T DISULFID 410 426 BY SIMILARITY.
T DISULFID 437 465 N-LINKED (GLNAC. . .) (POTENTIAL).
T CARBOHYD 213 213 N-LINKED (GLNAC. . .) (POTENTIAL).
T CARBOHYD 249 249 K -> N (IN DBSNP:1056602).
T VARIANT 449 449 /FTID=VAR_011692.
T MUTAGEN 255 R->Q: LOSS OF CLEAVAGE.
T MUTAGEN 441 441 S->A: LOSS OF ACTIVITY.
T CONFLICT 160 160 M -> V (IN REF. 3).
T CONFLICT 242 242 I -> L (IN REF. 1).
T CONFLICT 329 329 E -> Q (IN REF. 1).
T CONFLICT 489 491 RAD -> KAN (IN REF. 1).
T SEQUENCE 492 AA; 53891 MW; CAB44FD174A9076B CRC64;

Query Match 31.68; Score 506; DB 1; Length 492;
Best Local Similarity 43.88; Pred. No. 9.7e-40;
Matches 106; Conservative 37; Mismatches 85; Indels 14; Gaps 7;

Y 51 KIYGGVALDVSFPQWQSIQYDKQVCGSGILDPHWLTAHCFRKH-TDVFNNKVRAGS 109
:||||: :||||: :||||: :||||: :||||: :||||: :||||: :||||:
b 255 RIVGGSALPGAPWQVSLHVQNVHVCSSIIIPWIVTAHCFKVEKPLNNPFWHTAFAGI 314
:||||: :||||: :||||: :||||: :||||: :||||: :||||: :||||:

Y 110 DKLGSP----PSLAVAKIIIEFNPMY---PKNDIALMKLQPLTSGTVRPICLPFFD 162
:||||: :||||: :||||: :||||: :||||: :||||: :||||: :||||:
b 315 LR-QSFMFYGAGYQVEKVI---SHPNYDSKTKNDIALMKLQPLTENDLVKPVCLPNPG 370
:||||: :||||: :||||: :||||: :||||: :||||: :||||: :||||:

Y 163 EELTPATEPLIIGWFTKQNGKMSDILQASQVVIDSTFCNADDAQGVRTKMKCAGI 222
:||||: :||||: :||||: :||||: :||||: :||||: :||||: :||||:
b 371 MMQPEQLCWISGWGATEEK-GKTSVLNAKVLIIETQRCNSRYVYDNLITPAMICAGF 429
:||||: :||||: :||||: :||||: :||||: :||||: :||||: :||||:

Y 223 PEGGVDTCCGDSGGPLMYQSDQ-WHVVGIVSWGCGGPGSTPGVYTKVSAVLNWIYVWK 281
:||||: :||||: :||||: :||||: :||||: :||||: :||||: :||||:
b 430 LQGNVDSCQDSGGPLVTSKNNIWLIGDTSWGGCAKVRPGYGVNVVFTDWTIYQMR 489
:||||: :||||: :||||: :||||: :||||: :||||: :||||: :||||:

```

```

Oy 282 AE 283
Db 490 AD 491

RESULT 7
HEPS HUMAN
ID HEP5 HUMAN STANDARD; PRT; 417 AA.
AC P05981;
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-NOV-1988 (Rel. 09, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Serine protease hepsin (EC 3.4.21.-) (Transmembrane protease, serine
DE 1).
DE HFN OR TWPRSS1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=86209431; PubMed=2835076;
RA Leytus S.P., Loeb K.R., Hagen F.S., Kurachi K., Davie E.W.;
RT "A novel trypsin-like serine protease (hepsin) with a putative
transmembrane domain expressed by human liver and hepatoma cells.";
RL Biochemistry 27:1067-1074(1988).
[2]
RP SEQUENCE FROM N.A.
RC TISSUE=Pancreas, and Spleen;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Horg L.,
RA Scapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.J.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaby S.J.,
RA Richards S., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Villalón D., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalusz D.E.,
RA Schnurch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
[3]
RP CHARACTERIZATION.
RX MEDLINE=91358502; PubMed=1885621;
RA Tsuji A., Torres-Rosado A., Arai T., le Beau M.M., Lemons R.S.,
RA Chou S.H., Kurachi K.;
RT "Hepsin, a cell membrane-associated protease. Characterization,
tissue distribution, and gene localization.";
RL J. Biol. Chem. 266:16948-16953(1991).
[4]
RP CHARACTERIZATION.
RX MEDLINE=8346233; PubMed=8346233;
RA Torres-Rosado A., O'Shea K.S., Tsuji A., Chou S.H., Kurachi K.;
RT "Hepsin, a putative cell-surface serine protease, is required for
mammalian cell growth.";
CC -!- FUNCTION: Plays an essential role in cell growth and maintenance
of cell morphology.
CC -!- SUBCELLULAR LOCATION: Type II membrane protein.
CC -!- TISSUE SPECIFICITY: Present in most tissues, with the highest
level in liver.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.

```

 C This SWISS-PROT entry is copyright. It is produced through a collaboration
 C between the Swiss Institute of Bioinformatics and the EMBL outstation -
 C the European Bioinformatics Institute. There are no restrictions on its
 C use by non-profit institutions as long as its content is in no way
 C modified and this statement is not removed. Usage by and for commercial
 C entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 C or send an email to license@isb-sib.ch).
 C -----
 R EMBL; M18930; AAA36013.1; --
 R EMBL; X07732; CAA30558.1; --
 R EMBL; X07002; CAA30058.1; --
 R EMBL; BC025716; AAR25716.1; --
 R PIR; S00845; S00845.
 R HSP; P00763; IDPO.
 R MEROPS; S01.224; --
 R Genew; HGNC:5155; HPN.
 R MIM; 142440.
 R GO; GO:0005887; C: integral to plasma membrane; TAS.
 R GO; GO:0008151; P: cell growth and/or maintenance; TAS.
 R InterPro; IPR001314; Chymotrypsin.
 R Pfam; PF00089; trypsin_1.
 R PRINTS; PR00722; CHYMOTRYPsin.
 R SMART; SM00020; Tryp_Spc; 1.
 R PROSITE; PS0240; TRYPSIN_DOM; 1.
 R PROSITE; PS00134; TRYPSIN_HIS; 1.
 R PROSITE; PS00135; TRYPSIN_SER; 1.
 W Hydroxylase; Serine protease; Transmembrane; Signal-anchor.
 T CHAIN 1 162 SERINE PROTEASE; HEPsin, NON-CATALYTIC
 T CHAIN 163 417 CHAIN (POTENTIAL).
 T DOMAIN 1 17 SERINE PROTEASE HEPsin, CATALYTIC CHAIN
 T TRANSMEM 18 44 (POTENTIAL).
 T DOMAIN 45 417 CYTOPLASMIC (POTENTIAL).
 T DOMAIN 163 417 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
 T ACT_SITE 203 203 (POTENTIAL).
 T ACT_SITE 257 257 EXTRACELLULAR (POTENTIAL).
 T ACT_SITE 353 353 SERINE PROTEASE.
 T DISULFID 153 277 CHARGE RELAY SYSTEM (BY SIMILARITY).
 T DISULFID 188 204 CHARGE RELAY SYSTEM (BY SIMILARITY).
 T DISULFID 322 338 INTERCHAIN (BY SIMILARITY).
 T DISULFID 349 381 BY SIMILARITY.
 T CARBOHYD 112 112 BY SIMILARITY.
 Q SEQUENCE 417 AA; 45011 MW; B2086FF661E551D7 CRC64;
 Query Match 31.6%; Score 505.5; DB 1; Length 417;
 Best Local Similarity 41.1%; Pred. No. 8.9e-40;
 Matches 101; Conservative 36; Mismatches 94; Indels 15; Gaps 4;
 Y 50 DKIVGVYALVDVSPWQVSIQYDKQVHGVCGGSLDPHVLTAACFKRHTDVEN-WKVRAG 108
 b 161 DRIVGGRDTSIGRPFQWQSVRYDAHLICGSLSGDWLTHAHCFPERNVLSRWRFAG 220
 Y 109 SDKLGSFPFSLAVAKIIIEFNPMYP-----KNDIALMKLOPFLTFSGTVAPICLPF 160
 b 221 AVAQASPHGLQLGVQAVYVHGGLYFPDPNSPENSNDIALVHLSSPLPLTEYIQVCLPA 280
 Y 161 FDELTATPLWITGWFTKQNGKMSDILLOASQVVIDSTRCNADDAYQGEVTERKMC 220
 b 281 AGQALVPDKICTVTGWGNT-QYQQAQVGLQEARVPIISDNCVNGADFYGNQIKPRMFC 339
 Y 221 GIPRGGVDTCQDGGSGPLMYQ-----SDQWVGVGVSGVCGGPGSTPGVYTKYSAYLNW 275
 b 340 GYPGGGIDACQDGGGPFVCEDSISRTPRMELCGIVSMGTGCAALQKPGVYTKVSDPREW 399
 Y 276 IYNVWK 281
 b 400 IFQAIK 405

RESULT 8

ACRO MOUSE
 ID ACRO MOUSE STANDARD; PRT; 436 AA.
 DT 01-NOV-1991 (Rel. 20, Created)
 DT 01-NOV-1991 (Rel. 20, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Acrosin precursor (EC 3.4.21.10).
 GN ACR.
 OS Mus musculus (Mouse).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 CC NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-91185335; PubMed-2127931;
 RA Kashiwabara S., Baba T., Takada M., Watanabe K., Yano Y., Arai Y.;
 RT "Primary structure of mouse proacrosin deduced from the cDNA sequence
 RL and its gene expression during spermatogenesis.";
 RL J. Biochem. 108:785-791(1990).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-92041732; PubMed-1939002;
 RA Watanabe K., Baba T., Kashiwabara S., Okamoto A., Arai Y.;
 RT "Structure and organization of the mouse acrosin gene.";
 RL J. Biochem. 109:828-833(1991).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-90255839; PubMed-2111255;
 RA Klemm U., Maier W.-M., Teasouidou S., Adham I., Willison K.,
 RA Engel W.;
 RT "Mouse proacrosin: cDNA sequence, primary structure and
 RL postmeiotic expression in spermatogenesis.";
 RL Differentiation 42:160-166(1990).
 RN [4]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-92147126; PubMed-1783391;
 RA Kreamling H., Keime S., Wilhelm K., Adham I.M., Hameister H.,
 RA Engel W.;
 RT "Mouse proacrosin gene: nucleotide sequence, diploid expression, and
 RL chromosomal localization.";
 RL Genomics 11:828-834(1991).
 CC CC
 CC -1- FUNCTION: ACROSIN IS THE MAJOR PROTEASE OF MAMMALIAN SPERMATOZOA.
 CC IT IS A SERINE PROTEASE OF TRYPSIN-LIKE CLEAVAGE SPECIFICITY, IT
 CC IS SYNTHESIZED IN A ZYMOGEN FORM, PROACROSIN AND STORED IN THE
 CC ACROSOME.
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of Arg-|-Xaa and Lys-|-Xaa bonds;
 CC Preferential cleavage Arg-|-Xaa > Lys-|-Xaa > Lys-|-Xaa.
 CC -1- SUBUNIT: HEAVY CHAIN (CATALYTIC) AND A LIGHT CHAIN LINKED BY TWO
 CC DISULFIDE BONDS.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; S66245; AAB20293.1; --
 DR EMBL; S64500; AAB20293.1; JOINED.
 DR EMBL; S66243; AAB20293.1; JOINED.
 DR EMBL; D00754; BAA00651.1; ALT_INIT.
 DR EMBL; X52466; CAA36704.1; ALT_INIT.
 DR EMBL; M85170; AAA40124.1; ALT_INIT.
 DR EMBL; M96430; AAA37163.1; --
 DR EMBL; M96426; AAA37163.1; JOINED.
 DR EMBL; M96427; AAA37163.1; JOINED.
 DR EMBL; M96428; AAA37163.1; JOINED.
 DR PIR; JX0172; JX0172.
 DR HSP; P00763; IDPO.
 DR MEROPS; S01.223; --.

[illegible]


```
Y 173 IIGWGTQKQNGKMSDILLOASVOVIDSTRCNADDDAYQGEVTEKMCAGIPEGGVDTCCQ 232
b 180 VTGWGVIKONAPRSPVLMPEARVLDLDCNSTQWNGRVTSNVVAGYPEGKIDTCQ 239
Y 233 DSGGPLYMSQSD---WHVGVGVGVGVGGCGGPGVGVTKVSAVLNWI 276
b 240 DSGGPLYMCRDTRRQPFVIVGITSWGVGCARAKRPGVVTATWDYLDWI 286

RESULT 12
AL_RAT
D_KAL_RAT STANDARD; PRT; 638 AA.
T 01-JAN-1990 (Rel. 13, Created)
T 01-JAN-1990 (Rel. 13, Last sequence update)
T 28-FEB-2003 (Rel. 41, Last annotation update)
E Plasma kallikrein precursor (EC 3.4.21.34) (Plasma prekallikrein)
E (kininogenin) (flecher factor).
N KLK1 OR PK.
S Rattus norvegicus (Rat).
C Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
C Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
X NCBI_TaxID=101116;
[1]
SEQUENCE FROM N.A.
MEDLINE-91129236; PubMed=1993180;
A Beaubien G., Rosinski-Chupin I., Mattei M.-G., Mbikay M., Chretien M.,
A Seidah N.G.;
T "Gene structure and chromosomal localization of plasma kallikrein.";
L Biochemistry 30:1628-1635(1991).
[2]
SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
MEDLINE-90091743; PubMed=2598771;
A Seidah N.G., Ladenheim R., Mbikay M., Hamelin J., Lutfalla G.,
A Rougeon F., Lazure C., Chretien M.;
T "The cDNA structure of rat plasma kallikrein.";
L DNA 8:563-574(1989).
-!- FUNCTION: THE ENZYME CLEAVES LYS-ARG AND ARG-SER BONDS. IT
ACTIVATES, IN A RECIPROCAL REACTION, FACTOR XII AFTER ITS BINDING
TO A NEGATIVELY CHARGED SURFACE. IT ALSO RELEASES BRADYKININ FROM
HMW KININOGEN AND MAY ALSO PLAY A ROLE IN THE RENIN-ANGIOTENSIN
SYSTEM BY CONVERTING PRORENIN INTO RENIN.
-!- CATALYTIC ACTIVITY: cleaves selectively Arg-|-Xaa and Lys-|-Xaa
bonds, including Lys-|-Arg and Arg-|-Ser bonds in (human)
kininogen to release bradykinin.
-!- SUBUNIT: THE ZYMOGEN IS ACTIVATED BY FACTOR XIIA, WHICH CLEAVES
THE MOLECULE INTO A LIGHT CHAIN, WHICH CONTAINS THE ACTIVE SITE,
AND A HEAVY CHAIN, WHICH ASSOCIATES WITH HMW KININOGEN. THESE
CHAINS ARE LINKED BY ONE OR MORE DISULFIDE BONDS.
-!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1. PLASMA KALLIKREIN
SUBFAMILY.
-!- SIMILARITY: Contains 4 apple domains.
C This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
```

```
DR EMBL; M30282; AAA41463.1; -.
DR EMBL; M58590; AAA42069.1; -.
DR PIR; A39180; KORTPL.
DR KSSP; P00750; IRTF.
DR MEROPS; S01.212; -.
DR InterPro; IPR000177; Apple.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR003014; PAN.
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF00024; PAN; 4.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00005; APPLEDOMAIN.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00223; APPLE; 4.
DR SMART; SM00020; Tryp_Spc; 1.
DR PROSITE; PS00495; APPLE; 4.
DR PROSITE; PS0240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease; Glycoprotein; Plasma; Zymogen; Signal;
KW Fibrinolysis; Blood coagulation; Inflammatory response; Liver;
KW Repeat.
FT SIGNAL 1 19
FT CHAIN 20 390 PLASMA KALLIKREIN HEAVY CHAIN.
FT CHAIN 391 638 PLASMA KALLIKREIN LIGHT CHAIN.
FT DOMAIN 20 105 APPLE 1.
FT DOMAIN 110 195 APPLE 2.
FT DOMAIN 200 285 APPLE 3.
FT DOMAIN 291 376 APPLE 4.
FT DOMAIN 389 621 SERINE PROTEASE.
FT CARBOHYD 127 127 N-LINKED (GLCNAC. .) (PROBABLE).
FT CARBOHYD 308 308 N-LINKED (GLCNAC. .) (PROBABLE).
FT CARBOHYD 396 396 N-LINKED (GLCNAC. .) (PROBABLE).
FT CARBOHYD 453 453 N-LINKED (GLCNAC. .) (PROBABLE).
FT CARBOHYD 454 454 N-LINKED (GLCNAC. .) (PROBABLE).
FT ACT_SITE 434 434 CHARGE RELAY SYSTEM.
FT ACT_SITE 483 483 CHARGE RELAY SYSTEM.
FT ACT_SITE 578 578 CHARGE RELAY SYSTEM.
FT DISULFID 21 104 BY SIMILARITY.
FT DISULFID 47 77 BY SIMILARITY.
FT DISULFID 51 57 BY SIMILARITY.
FT DISULFID 111 194 BY SIMILARITY.
FT DISULFID 137 166 BY SIMILARITY.
FT DISULFID 141 147 BY SIMILARITY.
FT DISULFID 201 284 BY SIMILARITY.
FT DISULFID 227 256 BY SIMILARITY.
FT DISULFID 231 237 BY SIMILARITY.
FT DISULFID 292 375 BY SIMILARITY.
FT DISULFID 318 347 BY SIMILARITY.
FT DISULFID 322 328 BY SIMILARITY.
FT DISULFID 340 345 BY SIMILARITY.
FT DISULFID 383 503 BY SIMILARITY.
FT DISULFID 419 435 BY SIMILARITY.
FT DISULFID 517 584 BY SIMILARITY.
FT DISULFID 548 563 BY SIMILARITY.
FT DISULFID 574 602 BY SIMILARITY.
SQ SEQUENCE 638 AA; 71273 MW; 454BE27B8CA8F88 CRC64;
Query Match 30.6%; Score 489; DB 1; Length 638;
Best Local Similarity 42.2%; Pred. No. 5.2e-38;
Matches 100; Conservative 45; Mismatches 76; Indels 16; Gaps 8;
QY 51 KIVGGYALDVDSWPQVSIQ---YDKQHVCGSGILDPEHWLTAACPR--KHTDVFNKV 105
Db 390 RVVGTNSLGEWPMQVSLQVLSQNVNCGSGITGROWLITAAHCFDGIPIPDV--MRI 447
QY 106 RAG-----SDKLGSFPLSAVAKIIITFENPMYPKONDIALMKLQFPLTFSGTVRPICLPFF 161
Db 448 YGILNLSEITNKYPFSSIKELIHQKYMSEGSVDIALIKLQFLNTEYEQKPICLFSK 507
QY 162 DEELTPATPLMIIGWFTQKQNGKMSDILLOASVOVIDSTRCNADDDAYQGEV-TEKMC 220
Db 508 ADNTIYTNCTWGTGWTGER-GETQNILQKATIFLVNEEC--QKKYRDVYITQMICA 564
```

221 GIBGGVDTCCGDSGGPLMYO-SPQHHVVGIVSGYGGGSGSTPGVYTKVSAYLNIWI 276
 565 GYKGGIDACKDSGGPLVCKHSGRWLVGITSWGEGCARKEQPGVYTKVAEYIDWI 621

RESULT 13

MS-TW55_HUMAN STANDARD; PRT; 457 AA.
 Q9H383;
 16-OCT-2001 (Rel. 40, Created)
 16-OCT-2001 (Rel. 40, Last sequence update)
 28-FEB-2003 (Rel. 41, Last annotation update)
 Transmembrane protease, serine 5 (EC 3.4.21.-) (Spinesin).
 TMPS55.
 Homo sapiens (Human).
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 NCBI_TaxID=9606;
 [1]
 SEQUENCE FROM N.A.
 TISSUE=Brain;
 PubMed=11741986;
 Yamaguchi N., Okui A., Yamada T., Nakazato H., Mitsui S.;
 "Spinesin/TMPRSS5, a novel transmembrane serine protease, cloned from
 human spinal cord.";
 J. Biol. Chem. 277:6806-6812(2002).
 -1- SUBCELLULAR LOCATION: Type II membrane protein (Potential).
 -1- TISSUE SPECIFICITY: Brain-specific. Predominantly expressed in
 neurons, in their axons, and at the synapses of motoneurons in the
 spinal cord.
 -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
 -1- SIMILARITY: Contains 1 SRCR domain.
 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

 EMBL; AB028140; BAB20375.1; -
 HSSP; P00763; IDPO.
 Genew; HGNC:14908; TMPRSS5.
 MIM; 606751; -
 MEROPS; S01.313; -
 InterPro; IPR001314; Chymotrypsin.
 InterPro; IPR001254; Ser_protease_Try.
 InterPro; IPR001190; Src1_receptor.
 Pfam; PF00089; trypsin; 1.
 PRINTS; PR00722; CHYMOTRYPSIN.
 SMART; SM00020; Tryp_SPC; 1.
 PROSITE; PS50240; TRYPSIN_DOM; 1.
 PROSITE; PS00134; TRYPSIN_HIS; 1.
 PROSITE; PS00135; TRYPSIN_SER; 1.
 PROSITE; PS00420; SRCR_1; FALSE_NEG.
 PROSITE; PS50287; SRCR_2; FALSE_NEG.
 Hydrolyase; Serine protease; Transmembrane; Signal-anchor;
 Glycoprotein.
 DOMAIN 1 49 CYTOPLASMIC (POTENTIAL).
 TRANSMEM 50 70 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
 (POTENTIAL).
 DOMAIN 71 457 EXTRACELLULAR (POTENTIAL).
 DOMAIN 112 207 SRCR.
 DOMAIN 218 457 SERINE PROTEASE.
 ACT_SITE 258 258 CHARGE RELAY SYSTEM (BY SIMILARITY).
 ACT_SITE 308 308 CHARGE RELAY SYSTEM (BY SIMILARITY).
 ACT_SITE 405 405 CHARGE RELAY SYSTEM (BY SIMILARITY).
 SITE 217 218 CLEAVAGE (POTENTIAL).
 DISULFID 135 196 BY SIMILARITY.
 DISULFID 148 206 BY SIMILARITY.
 DISULFID 209 328 BY SIMILARITY.

FT DISULFID 243 259 BY SIMILARITY.
 FT DISULFID 374 390 BY SIMILARITY.
 FT DISULFID 401 429 BY SIMILARITY.
 FT CARBOHYD 163 163 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 170 170 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 170 170 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 195 195 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 319 319 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 375 375 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 457 AA; 49574 MW; 64406AB495A2651 CRC64;
 Query Match 30.4%; Score 487; DB 1; Length 457;
 Best Local Similarity 39.7%; Pred. No. 5.4e-38;
 Matches 94; Conservative 44; Mismatches 87; Indels 12; Gaps 5;
 QY 51 KIVGGVLDVDSWPQVSIQYDKQVHVGSGSILDPHVLTAAGCFR--KHTDVENWVKVRAG 108
 Db 217 RIVGGSVAPGRWPQASVALGFRHRCGGSVLAPRWVVTAAHCHMSFRLARLSSWRVHAG 276
 QY 109 SDKLGSFPLSLAVAKIIIEF---NPMYPKDN---DIALMKLQFPLTPSGTVTRICLPFFD 162
 Db 277 ---LVSHSAVRPHQGVVERIIIPLYSAQNHDYDVALLRLQALNFSDTVGAVCLPAKE 333
 QY 163 BEITPATPLWIIIGWGTIKONGKMSILLQASVQVIDSTFCNADDAYQGVETKMMCGAI 222
 Db 334 QHPKSGRCWVSGWGHTHPSHTYSSDMLQDTVVPLFSTQLCNSSCVYSGALTTPMLCAGY 393
 QY 223 PEGGVDTCCGDSGGPLMY-QSDQHHVVGIVSGYGGGSGSTPGVYTKVSAYLNIWI 278
 Db 394 LDGRADACQDSGGPLVCPDGTWRLVGVVSWGRACAEFNHPGVYAKVAEFLDWIHD 450
 RESULT 14
 TEST MOUSE STANDARD; PRT; 324 AA.
 AC Q9JH7; Q9DA14;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Testisin precursor (EC 3.4.21.-) (Trypsinase 4).
 GN PRSS21.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NCBI_TaxID=10090;
 RN [1]
 RC SEQUENCE FROM N.A.
 RP STRAIN=129/Sv.
 RX MEDLINE=21153229; PubMed=11231276;
 RA Scarman A.L., Hooper J.D., Boucaut K.J., Sit M.-L., Webb G.C.,
 RA Normyle J.F., Antalis T.M.;
 RA "Organization and chromosomal localization of the murine Testisin gene
 encoding a serine protease temporally expressed during
 spermatogenesis.";
 RL Eur. J. Biochem. 268:1250-1258(2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BALB/C; TISSUE=Testis;
 RX PubMed=11259427;
 RA Wong G.W., Li L., Madhusudan M.S., Krillis S.A., Gurish M.F.,
 RA Rothenberg M.E., Sali A., Stevens R.L.;
 RA "Trypsinase 4, a new member of the chromosome 17 family of mouse serine
 proteases.";
 RT J. Biol. Chem. 276:20648-20658(2001).
 RN [3]
 RP SEQUENCE OF 3-324 FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Testis;
 RX MEDLINE=21085660; PubMed=11217851;
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamahata I.,
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner T., Batalov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,


```

R EMBL; AF195508; AAF76458.1; -.
R EMBL; AF175759; AAF03697.1; -.
R EMBL; AF175522; AAF03695.1; -.
R EMBL; AF223563; AAG48852.2; -.
R HSP; P00763; IDPO.
R MEROPS; S01.028; -.
R Genew; HGNC:14134; TPG1.
R InterPro; IPR001314; Chymotrypsin.
R InterPro; IPR001254; Ser_protease_Try.
R Pfam; PF00089; trypsin; 1.
R PRINTS; PR00722; CHYMOTRYPSIN.
R PROSITE; PS0240; TRYPSIN_DOM; 1.
R PROSITE; PS0134; TRYPSIN_HIS; 1.
R PROSITE; PS00135; TRYPSIN_SER; FALSE NEG.
W Hydrolase; Serine protease; Signal; Glycoprotein; Zymogen;
W Transmembrane; Polymorphism.
T SIGNAL 1 19
T CHAIN 20 36
T CHAIN 38 321
T TRANSMEM 284 304
T ACT_SITE 78 78
T ACT_SITE 125 125
T ACT_SITE 222 222
T DISULFID 26 145
T DISULFID 63 79
T DISULFID 159 228
T DISULFID 192 210
T DISULFID 218 246
T CARBOHYD 85 85
T VARIANT 60 60
T VARIANT 126 126
T VARIANT 132 132
T VARIANT 204 204
T VARIANT 288 288
T CONFLICT 160 160
T SEQUENCE 321 AA; 33827 MW; FFF7B0E3C4A962D CRC64;

Query Match 30.0%; Score 480.5; DB 1; Length 321;
Best Local Similarity 40.6%; Pred.No. 1.4e-37;
Matches 97; Conservative 39; Mismatches 88; Indels 15; Gaps 7;

Y 47 DDDKIVGGYALDVDSWPFQVSIQYDKQHYCGSGILDPHWLTAACHFRKHTDVFNMKVR 106
b :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
33 DAGGRIVGGHAPAGAWFQASLRLRRMHVCGGSLSPQWVLTAAHCFSGSLNSSDYQVH 92
Y 107 AGSDKLGSPFLAVAKIIITEFNP--MYPKNDIALMKLOPFTFTGTVRPICLPFFDEE 164
b :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
93 LGELEITLSPHFSTVRIILHSSPSGQGTSGDIALVELSPVPTLSSRIPLPCLPEASDD 152
Y 165 LTPATPLWIIWGFTKONGKMGSD--ILLQASVQVIDSTRCNADDAVQGE---VTEKMM 218
b :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
153 FCGIRCWVTVGWYTR-GEPLPPPSLRVKVSVVDIETCRDD--YPPGGGSILOPDM 209
Y 219 CAGIPGGVDTCQGGSGGLMYQ--SDQWVHVGVWNGYCGGSPSTPGVYTKVSAYLNWI 276
b :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
210 CA---RPGGDACQDDSGGLVLCVQNGAWVQAGIVSWGEGCGRPNRPGVYTRVAYVNI 265

```

Search completed: November 23, 2003, 07:49:49
 DB time : 12.2462 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

M protein - protein search, using sw model

un on: November 23, 2003, 07:45:10 ; Search time 43.7799 Seconds
(without alignments)
1721.140 Million cell updates/sec

itle: US-09-607-745-9

erfect score: 1600

equences: 1 MDSKSSQKRLLLLVVSN.....LNWIYNVWKAELSRHHHHH 292

oring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

earched: 830525 seqs, 258052604 residues

otal number of hits satisfying chosen parameters: 830525

inimum DB seq length: 0

aximum DB seq length: 2000000000

ost-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

atabase :

SPTREMBL 23:*

1: sp_archaea:*

2: sp_bacteria:*

3: sp_fungi:*

4: sp_human:*

5: sp_invertebrate:*

6: sp_mammal:*

7: sp_mbc:*

8: sp_organelle:*

9: sp_phage:*

10: sp_plant:*

11: sp_rodent:*

12: sp_virus:*

13: sp_vertebrate:*

14: sp_unclassified:*

15: sp_virus:*

16: sp_bacteriapi:*

17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

result No.	Score	Query Match	Length	DB ID	Description
1	1266	79.1	405	4 Q96B86	Q96B86 homo sapien
2	1048	65.5	435	11 Q8VCA5	Q8VCA5 mus musculus
3	533	33.3	537	4 Q9BYE1	Q9BYE1 homo sapien
4	528	33.0	471	11 Q8CFE0	Q8CFE0 mus musculus
5	528	33.0	581	4 Q9BYE2	Q9BYE2 homo sapien
6	526.5	32.9	279	11 Q9QZ74	Q9QZ74 rattus norv
7	524	32.8	453	11 Q8VDE0	Q8VDE0 mus musculus
8	523	32.7	453	11 Q8K170	Q8K170 mus musculus
9	522	32.6	371	11 Q8CJ16	Q8CJ16 rattus norv
10	522	32.6	445	11 Q8CJ17	Q8CJ17 rattus norv
11	522	32.6	767	13 Q9DGR2	Q9DGR2 xenopus lae
12	521.5	32.6	417	11 Q8VHJ4	Q8VHJ4 rattus norv
13	519	32.4	455	11 Q8CDR0	Q8CDR0 mus musculus
14	509.5	31.8	417	11 Q8VDV1	Q8VDV1 mus musculus
15	509.5	31.8	417	11 Q8VHK8	Q8VHK8 mus musculus
16	506	31.6	492	4 Q96T73	Q96T73 homo sapien

17	492.5	30.8	799	11 Q9DBI0	Q9DBI0 mus musculus
18	487.5	30.5	329	6 Q9GLI0	Q9GLI0 ovis aries
19	486	30.4	638	11 Q8ROP5	Q8ROP5 mus musculus
20	483	30.2	490	11 Q92OK3	Q92OK3 rattus norv
21	481.5	30.1	377	6 P79343	P79343 bos taurus
22	480.5	30.0	423	11 Q8BML0	Q8BML0 mus musculus
23	478.5	29.9	321	4 Q96RZ8	Q96RZ8 homo sapien
24	478.5	29.9	421	11 Q60491	Q60491 cavia porce
25	477.5	29.8	416	11 Q8B230	Q8B230 mus musculus
26	477.5	29.8	416	11 Q8B213	Q8B213 mus musculus
27	477.5	29.8	802	4 Q8IUE2	Q8IUE2 homo sapien
28	477.5	29.8	811	4 Q8IU80	Q8IU80 homo sapien
29	473.5	29.6	415	6 Q29015	Q29015 sus sp. pre
30	470	29.4	1524	13 Q91674	Q91674 xenopus lae
31	469.5	29.3	813	11 Q9ROW3	Q9ROW3 rattus norv
32	468	29.2	643	6 Q97506	Q97506 sus scrofa
33	467	29.2	439	11 Q8BHM9	Q8BHM9 mus musculus
34	464.5	29.0	239	4 Q8NIC9	Q8NIC9 homo sapien
35	463	28.9	572	11 Q8BIK6	Q8BIK6 mus musculus
36	459.5	28.7	310	11 Q9QYZ9	Q9QYZ9 mus musculus
37	458	28.6	260	13 Q9W7P9	Q9W7P9 paralichthy
38	458	28.6	274	5 Q17086	Q17086 anopheles s
39	455.5	28.5	322	11 Q920S2	Q920S2 mus musculus
40	455	28.4	284	4 Q8NFR6	Q8NFR6 homo sapien
41	454.5	28.4	310	11 Q91XC4	Q91XC4 mus musculus
42	454	28.4	327	4 Q8N171	Q8N171 homo sapien
43	453	28.3	339	11 Q99L44	Q99L44 mus musculus
44	452.5	28.3	247	4 Q8NHM4	Q8NHM4 homo sapien
45	452.5	28.3	257	11 Q8BZ04	Q8BZ04 mus musculus

ALIGNMENTS

RESULT 1
Q96B86 PRELIMINARY; PRT; 405 AA.
ID Q96B86 AC Q96B86
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Similar to transmembrane protease, serine 4 (fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Pancreas;
RA Strausberg R.;
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
DR EMBL; BC012752; AAH12752.1; -
DR HSSP; P00761; 1AN1.
DR MEROPS; S01.034; -
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR002172; LDL receptor A.
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF00057; Srrc_receptor.
DR Pfam; PF00057; ldl_recept_a; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00192; LDLa; 1.
DR SMART; SM00202; SR; 1.
DR SMART; SM00020; Tryp_SPC; 1.
DR PROSITE; PS50287; SRRc_2; 1.
DR PROSITE; PS50240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SRR; 1.
KW Hydrolase; Protease; Serine protease; Transmembrane.
FT NON_TER 1
SEQUENCE 405 AA; 44474 MW; 951ACD52D9D48E04 CRC64;

```

Query Match      79.1%; Score 1266; DB 4; Length 405;
Best Local Similarity 97.4%; Pred. No. 4e-118;
Matches 228; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Y 51 KIVGGYALVDVSWPQVSIQYDKQHVCGGSIIDPHWLTAAHCFKHTDVFNNKVRAGSD 110
b 172 RVVGGEEASVDSWPQVSIQYDKQHVCGGSIIDPHWLTAAHCFKHTDVFNNKVRAGSD 231
Y 111 KLGSPFLSAVAKIIIEFNPMYPKNDIALMKLOPLTFSGTVRPICLPFFDEELTPATP 170
b 232 KLGSPFLSAVAKIIIEFNPMYPKNDIALMKLOPLTFSGTVRPICLPFFDEELTPATP 291
Y 171 LWIIGWFTKQNGKXSDILLQASVQVIDSTRCNADDAVQGEVTEKMCAGIPGGVDTC 230
b 292 LWIIGWFTKQNGKXSDILLQASVQVIDSTRCNADDAVQGEVTEKMCAGIPGGVDTC 351
Y 231 QGDSGGPLMYQSDQHVHVGVISWGYCGGPGSTPGVTKVSAYLNWIYNWKAE 284
b 352 QGDSGGPLMYQSDQHVHVGVISWGYCGGPGSTPGVTKVSAYLNWIYNWKAE 405

RESULT 2
8VCAS5 PRELIMINARY; PRT; 435 AA.
C Q8VCA5;
T 01-MAR-2002 (TREMELrel. 20, Created)
T 01-MAR-2002 (TREMELrel. 20, Last sequence update)
T 01-MAR-2003 (TREMELrel. 23, Last annotation update)
E Similar to transmembrane protease, serine 4 (Channel-activating
E protease 2).
N TMPSRS4.
S Mus musculus (Mouse).
C Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
C Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
X NCBI_taxID=10090;
N [1]
P SEQUENCE FROM N.A.
C TISSUE=Breast tumor;
A Strausberg R.;
L Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
N [2]
P SEQUENCE FROM N.A.
X MEDLINE=2214321; PubMed=12149280;
A Vauguiaux G., Vallet V., Jaeger N.F., Hummler E., Rossier B.C.;
T "Synergistic Activation of ENaC by Three Membrane-bound Channel-
T activating Serine Proteases (mCAP1, mCAP2, and mCAP3) and Serum- and
T Glucocorticoid-regulated Kinase (Sgk1) in Xenopus Oocytes.";
T J. Gen. Physiol. 120:131-201(2002).
L C -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
R EMBL; BC021368; AAH21368.1; -.
R EMBL; AY043240; AAK85307.1; -.
R HSSP; P00761; 1AN1.
R MEROPS; S01.034; -.
R MGD; MGLI2384877; Tmpres4.
R InterPro; IPR001314; Chymotrypsin.
R InterPro; IPR002172; LDL_receptor_A.
R InterPro; IPR001254; Ser_protease_Try.
R InterPro; IPR001190; S-cr_receptor.
R Pfam; PF00057; ldl_recept_a; 1.
R PRINTS; PR00722; Chymotrypsin.
R PROSITE; PS00089; trypsin_1.
R PROSITE; PS0287; SRCR_2; 1.
R PROSITE; PS0240; TRYPsin_DOM; 1.
R PROSITE; PS00134; TRYPsin_HIS; 1.
R PROSITE; PS00135; TRYPsin_SER; 1.
R Hydrolase; Protease; Serine protease; Transmembrane.
R SEQUENCE 435 AA; 47495 MW; DC52E45A43E01369 CRC64;

Query Match      65.5%; Score 1048; DB 11; Length 435;
Best Local Similarity 78.6%; Pred. No. 2.9e-96;
Matches 184; Conservative 27; Mismatches 23; Indels 0; Gaps 0;

Y 51 KIVGGYALVDVSWPQVSIQYDKQHVCGGSIIDPHWLTAAHCFKHTDVFNNKVRAGSD 110

```

```

Db 202 RVVGGVERPVDVSWPQVSIQYDKQHVCGGSIIDPHWLTAAHCFKHTDVFNNKVRAGSN 261
QY 111 KLGSPFLSAVAKIIIEFNPMYPKNDIALMKLOPLTFSGTVRPICLPFFDEELTPATP 170
Db 262 ILGNSPFLSAVAKIIIEFNPMYPKNDIALMKLOPLTFSGTVRPICLPFFDEELTPATP 321
QY 171 LWIIGWFTKQNGKXSDILLQASVQVIDSTRCNADDAVQGEVTEKMCAGIPGGVDTC 230
Db 322 VWIIGWFTKQNGKXSDILLQASVQVIDSTRCNADDAVQGEVTEKMCAGIPGGVDTC 381
QY 231 QGDSGGPLMYQSDQHVHVGVISWGYCGGPGSTPGVTKVSAYLNWIYNWKAE 284
Db 382 QGDSGGPLMYQSDQHVHVGVISWGYCGGPGSTPGVTKVSAYLNWIYNWKAE 435

RESULT 3
Q9BYE1 PRELIMINARY; PRT; 537 AA.
AC Q9BYE1;
DT 01-JUN-2001 (TREMELrel. 17, Created)
DT 01-JUN-2001 (TREMELrel. 17, Last sequence update)
DT 01-MAR-2003 (TREMELrel. 23, Last annotation update)
DE Mosaic serine protease.
GN MSPS.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_taxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Lung;
RX MEDLINE=21167393; PubMed=11267681;
RA Kim D.R., Sharmin S., Inoue M., Kido H.;
RT "Cloning and expression of novel mosaic serine proteases with and
RT without a transmembrane domain from human lung.";
RL Biochim. Biophys. Acta 1518:204-209(2001).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
DR EMBL; AB048797; BAB39742.1; -.
DR HSSP; P00763; IDPO.
DR MEROPS; S01.087; -.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR002172; LDL_receptor_A.
DR InterPro; IPR001254; Ser_protease_Try.
DR InterPro; IPR001190; S-cr_receptor.
DR Pfam; PF00057; ldl_recept_a; 1.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; Chymotrypsin.
DR SMART; SM00192; LDLa; 1.
DR SMART; SM00202; SR; 1.
DR SMART; SM00202; Tryp_Spc; 1.
DR PROSITE; PS0287; SRCR_2; 1.
DR PROSITE; PS0240; TRYPsin_DOM; 1.
DR PROSITE; PS00134; TRYPsin_HIS; 1.
DR PROSITE; PS00135; TRYPsin_SER; 1.
R Hydrolase; Protease; Serine protease.
R SEQUENCE 537 AA; 58102 MW; A39FF4E8816DAECF CRC64;

Query Match      33.3%; Score 533; DB 4; Length 537;
Best Local Similarity 44.0%; Pred. No. 1.4e-44;
Matches 106; Conservative 38; Mismatches 87; Indels 10; Gaps 5;

QY 51 KIVGGYALVDVSWPQVSIQYDKQHVCGGSIIDPHWLTAAHCF--RKHTDVFNNKVRAG 108
Db 295 RIYGGALASDKWPQVSLHFGTTHICGGLTDAQWLVTAHAAHCFVTVREKVLGKVVYAG 354
QY 109 SDKLGSPFLSA-VAKIIIEFNPMY---PKNDIALMKLOPLTFSGTVRPICLPFFDEE 164
Db 355 TSNLHQLPRAASIAEII---NSNYTDEEDDYIALMRSLKPLTLSAHHPACLPHVGT 411
QY 165 LTPATPLTWIGWGFTKQNGKXSDILLQASVQVIDSTRCNADDAVQGEVTEKMCAGIPE 224
Db 412 FSLNETCWITGFKTETDDKTSFPLREVQVNLIDFKKNDILVYDSYLTTPRMWAGDLH 471

```



```

Y 167 PATPLWIIIGWGTQKNGKMSDILLOASVQVVIDSTRCNADDAVQGEVTEKMMKAGIPBG 226
b 650 AGTTTWISGWSYTYE-GGSVSYLYQAAIPLDLSNVQSYVINGQITSSMTICAGYLSGG 708
Y 227 VDTCCGDSGGPLW-YQSDQHWVGVISWGYGCGGSPSTRGVTKVSAYLNIWYN 278
b 709 VDTCCGDSGGPLVKNKNGTWLVGVTSWGDGCARANKPGVIGNVTTFLEWIS 761

RESULT 12
bVHJ4 PRELIMINARY; PRT; 417 AA.
C O8VHJ4
D O8VHJ4
T 01-MAR-2002 (TrEMBLrel. 20, Created)
T 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
T 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
E Airway trypsin-like protease.
S Rattus norvegicus (Rat).
C Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
C Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
X NCBI_TaxID=10116;
N SEQUENCE FROM N.A.
P STRAIN=Wistar; TISSUE=Trachea;
C Hansen I.A., Fassnacht M., Hammer F., Schammann M., Allolio B.;
A "Cloning and characterization of RAT, the airway trypsin-like protease
T of Rattus norvegicus.";
T Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
L -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
C EMBL; AF453776; AAL50817.1; -.
R HSP; P00761; 1ANI.
R MEROPS; S01.047; -.
R InterPro; IPR001314; Chymotrypsin.
R InterPro; IPR000082; SEA domain.
R InterPro; IPR001254; Ser_protease_Try.
R Pfam; PF01390; SEA; 1.
R Pfam; PF00089; trypsin; 1.
R PRINTS; PR00722; CHYMOTRYPSIN.
R SMART; SM00200; SEA; 1.
R SMART; SM00020; Tryp_SPC; 1.
R PROSITE; PS50024; SEA; 1.
R PROSITE; PS50240; TRYPSIN_DOM; 1.
R PROSITE; PS00134; TRYPSIN_HIS; 1.
R PROSITE; PS00135; TRYPSIN_SER; 1.
W Hydroxase; Protease; Serine protease.
Q SEQUENCE 417 AA; 46287 MW; DB9504158B018E21 CRC64;

Query Match 32.6%; Score 521.5; DB 11; Length 417;
Best Local Similarity 42.1%; Pred. No. 1.4e-43;
Matches 98; Conservative 44; Mismatches 82; Indels 9; Gaps 5;
Y 49 DDKIVGGYALVDVSWPQVSIQYDKQHVCGGSIIDPHWVLTAAHCFRKHDTDFVNNKVRAG 108
b 183 EERIIGGTAQETGDPWQVSIQLNNVHCGGTLISNLVLTAAHCFRYSYSPQQTATFG 242
Y 109 SDKLGSPFLAVAKIILII--EFNPMYKPDNDIALMKQLFLTSGTGVTRPCLPFBDEL 165
b 243 VSTIS--PRLRVRRAILAAHAEYNSI-TRDNDIAVQLDRFVETFRNTHRVCLPAAQNI 299
Y 166 TPATPLWIIIGWGTQKNGKMSDILLOASVQVVIDSTRCNADDAVQGEVTEKMMKAGIPBG 225
b 300 MPDSVAVYTGWSLTYGNTVYN-LQQGEVRIVSSEVNEFAGYGGVLPQMLCAGVRSG 358
Y 226 GVDTCQDGGSLMYQSDQ--WHVVGVISWGYGCGGSPSTRGVTKVSAYLNIWYN 278
b 359 AVDACQDGGSLPQVEDTRFLWVGVISWGYGCGGLPNKPGVITFVAYRWI 411

RESULT 13
bQCDRO
D Q8CDRO
C Q8CDRO; PRELIMINARY; PRT; 455 AA.

```

```

DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DE 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Transmembrane protease.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Testis;
RX MEDLINE=22354883; PubMed=12466851;
RA The FANTOM Consortium,
RA "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RT Nature 420:563-573(2002).
RL EMBL; AK029714; BAC26577.1; -.
SQ SEQUENCE 455 AA; 49669 MW; BE22BE2E7503C74B CRC64;

Query Match 32.4%; Score 519; DB 11; Length 455;
Best Local Similarity 42.4%; Pred. No. 2.9e-43;
Matches 101; Conservative 41; Mismatches 82; Indels 14; Gaps 6;
QY 51 KIVGGYALVDVSWPQVSIQYDKQHVCGGSIIDPHWVLTAAHCFRKHDTDFVNNKVRAG 107
Db 217 RIVGGQAVASGRWPQASVMLGSRHTCGASVLAHPHVVTAAHCMYSFRL-SRLSSNRVHA 275
QY 108 GSDKLGSPFLAVAKIILII--EFNPMYKPDN---DIALMKQLFLTSGTGVTRPCLPFF 161
Db 276 GLVSHGAVRQHGVTWVEKILP---HPLYSQNHDYDALLQLRTPINFSDTVGAVCLPAK 332
QY 162 DEELTPATPLWIIIGWGTQKNGKMSDILLOASVQVVIDSTRCNADDAVQGEVTEKMMKAG 221
Db 333 EQHPFWSGQCVWSWGHTDPSHTSSSTLQDTWYPLSLTYLCSNCSMYSGALTHERMLCAG 392
QY 222 IPEGVTCQDGGSLMYQSDQ--WHVVGVISWGYGCGGSPSTRGVTKVSAYLNIWYN 278
Db 393 YLDGRADACQDGGSLPQVEDTRFLWVGVISWGYGCGAEPNRPQVYAKVAEFLDWIHD 450

RESULT 14
Q8VDVI PRELIMINARY; PRT; 417 AA.
AC Q8VDVI;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DB Similar to airway trypsin-like protease.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RA Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
L -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
C EMBL; BC020151; AAB20151.1; -.
DR HSP; P00761; 1ANI.
DR MEROPS; S01.047; -.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR000082; SEA domain.
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF01390; SEA; 1.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00200; SEA; 1.
DR SMART; SM00020; Tryp_SPC; 1.
DR PROSITE; PS50024; SEA; 1.
DR PROSITE; PS50240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.

```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

M protein - nucleic search, using frame_plus_p2n model

un on: November 23, 2003, 07:52:05 ; Search time 3687.55 Seconds
(without alignments)
3239.440 Million cell updates/sec

itle: US-09-607-745-9
erfect score: 1600
equene: 1 MDSKSSQKSRLLLLLVSN.....LNWIYVWKAELSRHHHHH 292

coring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

earched: 2888711 seqs, 2045481386 residues

otal number of hits satisfying chosen parameters: 5777422

inimum DB seq length: 0

aximum DB seq length: 2000000000

ost-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

ommand line parameters:

MODEL=frame+p2n.model -DEV=xlh
Q=/cgn2_1/USPTO.spool/US09607745/runat_21112003_144345_22184/app_query.fasta_1.1038
DB=GenEmbl -QWMT=fastap -SUFFIX=rge -MINMATCH=0 -LOOPECL=0 -LOOPEXT=0
UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
USER=US09607745@cgn_1_1_5308 @runat_21112003_144345_22184 -NCPU=6 -ICPU=3
NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAEXT=0.5 -FGAPOP=6
FGAEXT=7 -YGAPOP=10 -YGAEXT=0.5 -DELOP=6 -DELEXT=7

atabase :

GenEmbl.*

1: gb_ba.*
2: gb_htg.*
3: gb_in.*
4: gb_lm.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vi.*
15: em_ba.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_mu.*
20: em_or.*
21: em_ov.*
22: em_ov.*
23: em_pat.*
24: em_ph.*
25: em_pl.*
26: em_ro.*
27: em_sts.*
28: em_un.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	1266	79.1	1305	6	AX076190 Sequence
2	1266	79.1	1793	9	BC012752 Homo sapi
3	1266	79.1	2038	6	AR142620 Sequence
4	1266	79.1	2038	6	BD137129 Human pro
5	1266	79.1	2063	6	AX092380 Sequence
6	1266	79.1	2063	6	AX376262 Sequence
7	1266	79.1	2063	6	AX395214 Sequence
8	1266	79.1	2063	6	AX697206 Sequence
9	1266	79.1	2079	6	AR233520 Sequence
10	1266	79.1	2079	6	AX207967 Sequence
11	1266	79.1	2079	9	AF216312 Homo sapi
12	1266	79.1	2081	9	AF179224 Homo sapi
13	1266	79.1	2088	9	BC011703 Homo sapi
14	1264	79.0	699	6	E13203 Human cDNA
15	1253	78.3	1479	6	AX076192 Sequence
16	1248	78.0	2137	6	AX207899 Sequence
17	1050	65.6	2060	10	BC021368 Mus muscu
18	1050	65.6	2268	10	AY043240 Mus muscu
19	765	47.8	1130	6	AR256996 Sequence
20	754.5	47.2	1130	6	AR234337 Sequence
21	746	46.6	1169	6	AR219284 Sequence
22	740.5	46.3	1166	6	AR221273 Sequence
23	672.5	42.0	1052	6	AR219287 Sequence
24	645	40.3	1049	6	AR219286 Sequence
25	601.5	37.6	1103	6	AR219328 Sequence
26	593	37.1	1142	6	AR219285 Sequence
27	574	35.9	1037	6	AR219329 Sequence
28	546.5	34.2	1381	6	BD141852 Airway sp
29	539	33.7	2067	6	AX149581 Sequence
30	538	33.6	1555	6	BD141849 Airway sp
31	534.5	33.4	1538	6	BD141855 Airway sp
32	533	33.3	1078	6	AX190545 Sequence
33	533	33.3	1314	6	AX190543 Sequence
34	533	33.3	1314	6	AX190546 Sequence
35	533	33.3	1614	6	AX360089 Sequence
36	533	33.3	1689	6	AX149579 Sequence
37	533	33.3	1748	6	AX369323 Sequence
38	533	33.3	3222	9	BC018715 Homo sapi
39	533	33.3	3324	9	AB048797 Homo sapi
40	532	33.2	2237	9	AY190317 Homo sapi
41	528	33.0	2393	9	AB048796 Homo sapi
42	528	33.0	2992	10	BC042878 Mus muscu
43	527.5	33.0	1576	6	BD141854 Airway sp
44	527	32.9	1045	10	AF198087 Rattus no
45	526.5	32.9	1460	6	AR225450 Sequence

ALIGNMENTS

RESULT 1

QY	191	LeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAspAlaTyrGln	210
DB	1024	CTGACGGCGTCAGTCAGGTCATTCACAGCACAGGGTCAATGCAGACGATGGTACCAG	1083
QY	211	GlyGluValThrGluLysMetMetCysalaGlyIleProGluGlyGlyValAspThrCys	230
DB	1084	GGGGAGTACCGAGBAGATGATGTGTGACAGCATCCCGAAGGGGTGTGGACACTGC	1143
QY	231	GlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGlyIle	250
DB	1144	CAGGTCACAGTGGTGGGCCCTGATGATGACCAATCTCACCAGTGGCATGTGGTGGGCATC	1203
QY	251	ValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyValTyrThrLysValSer	270
DB	1204	GTTAGCTGGGCTATGGCTGGGGGGCCCGAGCACCCACAGAGTATACACCAAGGTTCTCA	1263
QY	271	AlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu	284
DB	1264	GCCTATCTCAACTGGATCTACCAATCTCTGGAAGGTGAGCTG	1305
RESULT 2			
LOCUS	BC012752		
DEFINITION	Homo sapiens, Similar to transmembrane protease, serine 4, clone IMAGE:3623466, mRNA, partial cds.		
ACCESSION	BC012752		
VERSION	BC012752.1	GI:15215322	
KEYWORDS	Homo sapiens (human)		
SOURCE	Homo sapiens		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
AUTHORS	1 (bases 1 to 1793)		
TITLE	Strausberg, R.		
JOURNAL	Direct Submission		
REMARK	Submitted (15-AUG-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA		
COMMENT	NIH-MGC Project URL: http://mgc.nci.nih.gov Contact: MGC help desk Email: cgapps@email.nih.gov Tissue Procurement: ARCC cDNA Library Preparation: Rubin Laboratory cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Institute for Systems Biology http://www.systemsbio.org contact: amadansystemsbiology.org Anup Madan, Rachel Dickhoff, Jessica Fahey, Stephanie Ford, Julia Greene, Mark Kettelman and Anuradha Madan		
FEATURES	<p>Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Series: IRAL Plate; 24 Row; m Column: 19. Location/Qualifiers</p> <p>1..1793</p> <p>/organism="Homo sapiens"</p> <p>/mol_type="mRNA"</p> <p>/db_xref="taxon:9606"</p> <p>/clone="IMAGE:3623466"</p> <p>/tissue_type="Pancreas, adenocarcinoma"</p> <p>/clone_lib="NIH MGC 39"</p> <p>/lab_host="DH10B-R"</p> <p>/notes="vector: pOTe7"</p> <p><1..1219</p> <p>/codon_start=2</p> <p>/product="Similar to transmembrane protease, serine 4"</p> <p>/protein_id="AA512752.1"</p> <p>/db_xref="GI:15215323"</p>		
CDS			

CLACGKSLKTRPVGGEASVDSWPMQSVIOYDKOHVCGSILDPHVLTAACPRKH
TIVFNWKYRAGSKLSPSLAVAKLIIIEFNPMPKNDIALMKIQFLTFSGVRP
ICLPPDESLTATPILWIIIGWFTKQNGKMSDILQASVQVIDTRCNADDAYQGEV
TERKWCAGIPGEGVDTCQSDSGPLMYQSDQHVHVGIVSWGYCGGSPSTPGVYTKVSA
YLNWYNWKAEL"

BASE COUNT 436 a 501 c 486 g 370 t

ORIGIN

Alignment Scores:

red. No.: 9.65e-122 Length: 1793
Score: 1266.00 Matches: 228
Percent Similarity: 98.29% Conservative: 2
Best Local Similarity: 97.44% Mismatches: 4
Query Match: 79.12% Indels: 0
Gaps: 9

US-09-607-745-9 (1-292) x BC012752 (1-1793)

51 LysileValGlyGlyTyralaLeuAspValaspSerTrrProTrrGlnValSerileGln 70
b 515 COTGTGGTGGGTGGGAGGAGGCTCTGTGGATCTTGGCCCTGGCAGGTCCAGTCCAG 574
71 TyrAspLysGlnHisValCysGlyGlySerileLeuAspProHisTrrValLeuThra 90
b 575 TAGGACAAACAGACAGTCTGTGGAGGAGCATCTGTGACCCCATCTGGGTCTCAGCGCA 634
91 AlaHisCysPheArgLysHisThrAspValPheAsnTrrLysValargAlaGlySerAsp 110
b 635 GCCCACTGCTTCAGGAACAATACCGATGTGTCACTGGAAGGTGGCGAGCTCAGAC 694
111 LysLeuGlySerPheProSerLeuAlaValAlaLysileileileileGluPheAsnPro 130
b 695 AAACCTGGGAGCTTCCCATCTGCTGGTGTGGCCAGATCATCATCATTAATCAACCC 754
131 MetTrrProLysAspAsnAspileAlaLeuMetLysLeuGlnPheProLeuThrPheSer 150
b 755 ATGTACCCCAAGACAAATGACATCGCCCTCATGAGCTGACGTCCCATCTCACTTCTCA 814
151 GlyThrValArgProLysCysLeuProPhePheAspGluLeuThrProAlaThrPro 170
b 815 GGCACAGTCAGGCCCATCTGTGCTGCTTCTTGTGAGGAGTCACTCCAGCCACCCCA 874
171 LeuTrrPheileGlyTrrGlyPheThrLysGlnAsnGlyGlyLysMetSerAspileLeu 190
b 875 CTCTGGATCATTTGGATGGGGCTTTAGAAAGCAATGGAGGAAGATGCTCAGCATACTG 934
191 LeuGlnAlaSerValGlnValileAspSerThrArgCysAsnAlaAspAlaTrrGln 210
b 935 CTGACAGGCTCAGTCCAGTCAATTGACACACACAGGTGCAATGCAGACGATCGTACCAG 994
211 GlyGluValThrGluLysMetMetCysAlaGlyLileProGluGlyGlyValAspThrCys 230
b 995 GGGGAAGTCACCGAAGATGATGTGTGAGGATCCCGAAGGGGGTGTGACACCTGTC 1054
231 GlnGlyAspSerGlyGlyProLeuMetTrrGlnSerAspGlnTrrPheValValGlyLile 250
b 1055 CAGGTGTACAGTGGTGGGGCTTGAATGATGATGATGATGATGATGATGATGATGATG 1114
251 ValSerTrrPglyTrrGlyCysGlyGlyProSerThrProGlyValTrrLysValSer 270
b 1115 GTTAGCTGGGCTATGGTCTGGGGGGCCCGAGACCCCGAGAGATATACCCAGGTCTCA 1174
271 AlaTrrLeuAsnTrrPheTrrAsnValTrrLysAlaGluLeu 284
b 1175 GCCTATCTCACTGATCTACAATCTGTGGAAGGCTGAGCTG 1216

RESULT 3

R142620

OCUS

EFINITION

Sequence 18 from patent US 6203979.

CCESION

AR142620

ERSTON

AR142620.1 GI:15103906

EYWORDS

SOURCE Unknown.

ORGANISM Unknown.

REFERENCE 1 (bases 1 to 2038)

AUTHORS Bandman,O., Hillman,J.L., Yue,H., Guegler,K.J., Corley,N.C.,

Tang,Y.Tom. and Shah,P.

TITLE Human protease molecules

JOURNAL Patent: US 6203979-A 18 20-MAR-2001;

FEATURES Location/Qualifiers

1..2038

source /organism="unknown"

BASE COUNT 462 a 591 c 569 g 416 t

ORIGIN

Alignment Scores:

Pred. No.: 1.13e-121 Length: 2038
Score: 1266.00 Matches: 228
Percent Similarity: 98.29% Conservative: 2
Best Local Similarity: 97.44% Mismatches: 4
Query Match: 79.12% Indels: 0
Gaps: 6

US-09-607-745-9 (1-292) x AR142620 (1-2038)

51 LysileValGlyGlyTyralaLeuAspValaspSerTrrProTrrGlnValSerileGln 70
b 803 COTGTGGTGGGTGGGAGGAGGCTCTGTGGATCTTGGCCCTGGCAGGTCCAGTCCAG 862
71 TyrAspLysGlnHisValCysGlyGlySerileLeuAspProHisTrrValLeuThra 90
b 863 TAGGACAAACAGACAGTCTGTGGAGGAGCATCTGTGACCCCATCTGGGTCTCAGCGCA 922
91 AlaHisCysPheArgLysHisThrAspValPheAsnTrrLysValargAlaGlySerAsp 110
b 923 GCCCACTGCTTCAGGAACAATACCGATGTGTCACTGGAAGGTGGCGAGCTCAGAC 982
111 LysLeuGlySerPheProSerLeuAlaValAlaLysileileileileGluPheAsnPro 130
b 983 AAACCTGGGAGCTTCCCATCTGCTGGTGTGGCCAGATCATCATTAATCAACCC 1042
131 MetTrrProLysAspAsnAspileAlaLeuMetLysLeuGlnPheProLeuThrPheSer 150
b 1043 ATGTACCCCAAGACAAATGACATCGCCCTCATGAGCTGACGTCCCATCTCACTTCTCA 1102
151 GlyThrValArgProLysCysLeuProPhePheAspGluLeuThrProAlaThrPro 170
b 1103 GGCACAGTCAGGCCCATCTGTGCTGCTTCTTGTGAGGAGTCACTCCAGCCACCCCA 1162
171 LeuTrrPheileGlyTrrGlyPheThrLysGlnAsnGlyGlyLysMetSerAspileLeu 190
b 1163 CTCTGGATCATTTGGATGGGGCTTTAGCAAGCAAGATGGAGGGAAGATGCTGACATACTG 1222
191 LeuGlnAlaSerValGlnValileAspSerThrArgCysAsnAlaAspAlaTrrGln 210
b 1223 CTGACAGGCTCAGTCCAGTCAATTGACACACACAGGTGCAATGCAGACGATGCGTACCAG 1282
211 GlyGluValThrGluLysMetMetCysAlaGlyLileProGluGlyGlyValAspThrCys 230
b 1283 GGGGAAGTCACCGAAGATGATGTGTGAGGATCCCGAAGGGGGTGTGACACCTGTC 1342
231 GlnGlyAspSerGlyGlyProLeuMetTrrGlnSerAspGlnTrrPheValValGlyLile 250
b 1343 CAGGTGTACAGTGGTGGGGCTTGAATGATGATGATGATGATGATGATGATGATGATG 1402
251 ValSerTrrPglyTrrGlyCysGlyGlyProSerThrProGlyValTrrLysValSer 270
b 1403 GTTAGCTGGGCTATGGTCTGGGGGGCCCGAGACCCCGAGAGATATACCCAGGTCTCA 1462
271 AlaTrrLeuAsnTrrPheTrrAsnValTrrLysAlaGluLeu 284
b 1463 GCCTATCTCACTGATCTACAATCTGTGGAAGGCTGAGCTG 1504

RESULT 4

171	LeuThrPillelleGlyTyrGlyPheThrLysGlnAsnGlyGlyLysMetSerAspIleLeu	190
1163	CTCTGGATCAATTGGATGGGCTTTACGAAGCAGAAATGAGGAAGATGTCTGACATAC	1222
191	LeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAlaTyrGln	210
1223	CTGCAGGCGTCAGTCCAGGTCATTGACAGCACACGGTGCAATCCAGACGATGGGTAC	1282
211	GlyGlnValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThrCys	230
1293	GGGGAAGTCACCGAGAAGATGATGTGCGAGGCATCCCGAGGGGTGTGGACCTGC	1342
231	GlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGlyIle	250
1343	CAGGTCACAGTCGTGGGCCCCCTGATGACCAATCTGACAGTGGCATGTGGTGGGCATC	1402
251	ValSerTyrGlyTyrGlyCysGlyGlyProSerThrProGlyValTyrThrLysValSer	270
1403	GTTAGCTGGGCTATTGGCTCGGGGGCCCCAGACCCCGAGAGTATACCAAGTCTCA	1462
271	AlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu	284
1463	GCTATCTCAACTGCATCTACATGTCTGGAAGGCTGAGCTG	1504
RESULT 5		
AX092380	AX092380	2063 bp
LOCUS	Sequence 111 from Patent WO0116318.	linear
DEFINITION	AX092380	
ACCESSION	AX092380.1	
VERSION	GI:13444504	
KEYWORDS	Homo sapiens (human)	
SOURCE	Homo sapiens (human)	
ORGANISM	Homo sapiens	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
AUTHORS	Eaton,D.L., Filvaroff,E., Gerritsen,M.E., Goddard,A., Godowski,P.J., Grimaldi,C.J., Gurney,A.L., Watanabe,C.K. and Wood,W.I.	
TITLE	Secreted and transmembrane polypeptides and nucleic acids encoding the same	
JOURNAL	Patent: WO 0116318-A 111 08-MAR-2001; Genentech, Inc. (US)	
FEATURES	Location/Qualifiers	
source	1..2063	
	/organism="Homo sapiens"	
	/mol_type="genomic DNA"	
	/db_xref="taxon:9606"	
BASE COUNT	477 a 591 c 576 g 419 t	
ORIGIN		
Alignment Scores:		
Pred. No.:	1.15e-121	Length: 2063
Score:	1266.00	Matches: 228
Percent Similarity:	98.29%	Conservative: 2
Best Local Similarity:	97.44%	Mismatches: 4
Query Match:	79.12%	Indels: 0
DB:	6	Gaps: 0
US-09-607-745-9 (1-292) x AX092380 (1-2063)		
QY	51 LysIleValGlyGlyTyrAlaLeuAspValAspSerTrpProTrpGlnValSerIleGln	70
Db	
804	CGTGTGGTGGGTGGGAGGAGCCCTCTGTGGATTCCTTGGCCCTTGGCAGGTCAGATCCAG	863
QY	71 TyrAspLysGlnHisValCysGlySerIleLeuAspProHisTrpValLeuThrAla	90
Db	
864	TACGACAAACAGCAGCTCTGTGGAGGGAGGCATCTCTGGACCCCTCTCCACGGCA	923
QY	91 AlatiHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySerAsp	110
Db	
924	GCCCACTCTCTTCAGGAACATACCGATGTCTCACTGGAAAGTGGGGAGGCTCAGAC	983
QY	111 LysLeuGlySerPheProSerLeuAlaValAlaLysIleIleLeuGluPheAsnPro	130
Db	
983	AAACTGGGCAGCTTCCCATCTCTGGCTGTGGCCAGATCATCATTTGATTCACCC	1042
QY	131 MetTyrProLysAspAsnAlaLeuMetLysLeuGlnPheProLeuThrPheSer	150
Db	
1043	ATGTACCCCAAGAACAATGACATCGCCCTCATGAAGCTGCAGTTCCTCCACTCTCTCA	1102
QY	151 GlyThrValArgProIleCysLeuProPhePheAspGluGluLeuThrProAlaThrPro	170
Db	
1103	GCACAGTCAGGCCCATCTGTCTGCCCTTCTTTGATGAGGAGCTCACTCCAGCCACCCA	1162
QY	171 LeuThrPillelleGlyTyrGlyPheThrLysGlnAsnGlyGlyLysMetSerAspIleLeu	190
Db	1163 CTCTGGATCAATTGGATGGGCTTTACGAAGCAGAAATGAGGAAGATGTCTGACATAC	1222
QY	191 LeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAlaTyrGln	210
Db	1223 CTGCAGGCGTCAGTCCAGGTCATTGACAGCACACGGTGCAATCCAGACGATGGGTAC	1282
QY	211 GlyGlnValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThrCys	230
Db	1293 GGGGAAGTCACCGAGAAGATGATGTGCGAGGCATCCCGAGGGGTGTGGACCTGC	1342
QY	231 GlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGlyIle	250
Db	1343 CAGGTCACAGTCGTGGGCCCCCTGATGACCAATCTGACAGTGGCATGTGGTGGGCATC	1402
QY	251 ValSerTyrGlyTyrGlyCysGlyGlyProSerThrProGlyValTyrThrLysValSer	270
Db	1403 GTTAGCTGGGCTATTGGCTCGGGGGCCCCAGACCCCGAGAGTATACCAAGTCTCA	1462
QY	271 AlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu	284
Db	1463 GCTATCTCAACTGCATCTACATGTCTGGAAGGCTGAGCTG	1504
RESULT 5		
AX092380	AX092380	2063 bp
LOCUS	Sequence 111 from Patent WO0116318.	linear
DEFINITION	AX092380	
ACCESSION	AX092380.1	
VERSION	GI:13444504	
KEYWORDS	Homo sapiens (human)	
SOURCE	Homo sapiens (human)	
ORGANISM	Homo sapiens	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
AUTHORS	Eaton,D.L., Filvaroff,E., Gerritsen,M.E., Goddard,A., Godowski,P.J., Grimaldi,C.J., Gurney,A.L., Watanabe,C.K. and Wood,W.I.	
TITLE	Secreted and transmembrane polypeptides and nucleic acids encoding the same	
JOURNAL	Patent: WO 0116318-A 111 08-MAR-2001; Genentech, Inc. (US)	
FEATURES	Location/Qualifiers	
source	1..2063	
	/organism="Homo sapiens"	
	/mol_type="genomic DNA"	
	/db_xref="taxon:9606"	
BASE COUNT	462 a 591 c 569 g 416 t	
ORIGIN		
Alignment Scores:		
Pred. No.:	1.13e-121	Length: 2038
Score:	1266.00	Matches: 228
Percent Similarity:	98.29%	Conservative: 2
Best Local Similarity:	97.44%	Mismatches: 4
Query Match:	79.12%	Indels: 0
DB:	6	Gaps: 0
US-09-607-745-9 (1-292) x BD137129 (1-2038)		
QY	51 LysIleValGlyGlyTyrAlaLeuAspValAspSerTrpProTrpGlnValSerIleGln	70
Db	</

Alignment Scores:

red. No.: 1.15e-121 Length: 2063
 core: 1266.00 Matches: 228
 Percent Similarity: 98.23% Conservative: 2
 Best Local Similarity: 97.44% Mismatches: 4
 Query Match: 79.12% Indels: 0
 DB: 6 Gaps: 0

S-09-607-745-9 (1-292) x AX395214 (1-2063)

Y 51 LysileValGlyGlyTyrAlaLeuAspValAspSerTrpProTrpGlnValSerIleGln 70
 b 804 CGTGTGGTGGTGGAGAGGCGCTCTGTGGATTCTTGGCCCTGGCAGTCCAGTCACG 863
 Y 71 TyrAspLysGlnHisValCysGlySerIleLeuAspProHisTrpValLeuThrAla 90
 b 864 TACGACAAACAGCAGCTCTGTGGAGGAGCAGTCCTGGACCCCACTGGGTCTCCAGGCA 923
 Y 91 AlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySerAsp 110
 b 924 GCCCACTGCTTCAGAAACATACCGATGTGTTCAACTGGAGGTGGGGCAGGCTCAGAC 983
 Y 111 LysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleIleIleIleIle 130
 b 984 AAACCTGGGAGCTTCCATCCCTGGCTGTGGCCAGATCATCATGATTGAATCAACCCC 1043
 Y 131 MetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPheSer 150
 b 1044 ATGTACCCCAAGACAAATGACATGCGCCCTCATGAGCTGACGTTCCCACTCACTTCTCA 1103
 Y 151 GlyThrValArgProIleCysLeuProPhePheAspGluGluLeuThrProAlaThrPro 170
 b 1104 GGCACAGTCAGGCCCATCTGTCTGCTCTCTTGTATGAGGAGCTCACTCCAGCCACCCCA 1163
 Y 171 LeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyGlyLysMetSerAspIleLeu 190
 b 1164 CTCGTGATCATGTGATGGGCTTTACAGACAGATGAGGGAAGATGCTGACATACCTG 1223
 Y 191 LeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAlaThrGln 210
 b 1224 CTGACGGCGTCAGTCCAGTCAATTGACAGCACACGCTGCAATGACAGATGCTGACACG 1283
 Y 211 GlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThrCys 230
 b 1284 GGGGAAGTACCCAGAGAAGATGATGTGTGAGCATCCCGAGAGGGGGTGTGACACCTGC 1343
 Y 231 GlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGlyIle 250
 b 1344 CAGGTGACAGTGTGGGCCCCCTGATGATACCAATCTGACAGTGGCATGTGTGGGCGATC 1403
 Y 251 ValSerTrpGlyTyrGlyCysGlyProSerThrProGlyValTyrThrLysValSer 270
 b 1404 GTTAGCTGGGCTATGCTGCGGGGCGGAGCAGCCCGAGAGGTATACACCAAGTCTCA 1463
 Y 271 AlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu 284
 b 1464 GCCTATCTCAACTGGATCTACATGCTGTGAGAGGTGAGCTG 1505

RESULT 8

X697206

LOCUS

DEFINITION

SEQUENCE 274 from Patent WO0078961.

AX697206

AX697206.1 GI:29498146

KEYWORDS

Hom sapiens (human)

ORGANISM

REFERENCE

AUTHORS

Ferrara, N., Stewart, T.A., Williams, P.M., Baker, K.P., Desnoyers, L.,

Eaton, D.L., Gao, W.Q., Pan, J., Botstein, D., Pong, S., Goddard, A.,

Godowski, P.J., Gurney, A.L., Smith, V., Tumas, D., Wood, W.I.,

TITLE

JOURNAL

Genentech Inc. (US)

location/Qualifiers

1. .2063

/organism="Homo sapiens"

/mol_type="genomic DNA"

/db_xref="taxon:9606"

BASE COUNT 477 a 591 c 576 g 419 t

ORIGIN

Alignment Scores:

Pred. No.: 1.15e-121 Length: 2063

Score: 1266.00 Matches: 228

Percent Similarity: 98.23% Conservative: 2

Best Local Similarity: 97.44% Mismatches: 4

Query Match: 79.12% Indels: 0

DB: 6 Gaps: 0

US-09-607-745-9 (1-292) x AX697206 (1-2063)

QY 51 LysileValGlyGlyTyrAlaLeuAspValAspSerTrpProTrpGlnValSerIleGln 70

DB 804 CGTGTGGTGGTGGAGAGGCGCTCTGTGGATTCTTGGCCCTGGCAGTCCAGTCACG 863

QY 71 TyrAspLysGlnHisValCysGlySerIleLeuAspProHisTrpValLeuThrAla 90

DB 864 TACGACAAACAGCAGCTCTGTGGAGGAGCAGTCCTGGACCCCACTGGGTCTCCAGGCA 923

QY 91 AlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySerAsp 110

DB 924 GCCCACTGCTTCAGAAACATACCGATGTGTTCAACTGGAGGTGGGGCAGGCTCAGAC 983

QY 111 LysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleIleIleIleIle 130

DB 984 AAACCTGGGAGCTTCCATCCCTGGCTGTGGCCAGATCATCATGATTGAATCAACCCC 1043

QY 131 MetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPheSer 150

DB 1044 ATGTACCCCAAGACAAATGACATGCGCCCTCATGAGCTGACGTTCCCACTCACTTCTCA 1103

QY 151 GlyThrValArgProIleCysLeuProPhePheAspGluGluLeuThrProAlaThrPro 170

DB 1104 GGCACAGTCAGGCCCATCTGTCTGCTCTCTTGTATGAGGAGCTCACTCCAGCCACCCCA 1163

QY 171 LeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyGlyLysMetSerAspIleLeu 190

DB 1164 CTCGTGATCATGTGATGGGCTTTACAGACAGATGAGGGAAGATGCTGACATACCTG 1223

QY 191 LeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAlaThrGln 210

DB 1224 CTGACGGCGTCAGTCCAGTCAATTGACAGCACACGCTGCAATGACAGATGCTGACACG 1283

QY 211 GlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThrCys 230

DB 1284 GGGGAAGTACCCAGAGAAGATGATGTGTGAGCATCCCGAGAGGGGGTGTGACACCTGC 1343

QY 231 GlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGlyIle 250

DB 1344 CAGGTGACAGTGTGGGCCCCCTGATGATACCAATCTGACAGTGGCATGTGTGGGCGATC 1403

QY 251 ValSerTrpGlyTyrGlyCysGlyProSerThrProGlyValTyrThrLysValSer 270

DB 1404 GTTAGCTGGGCTATGCTGCGGGGCGGAGCAGCCCGAGAGGTATACACCAAGTCTCA 1463

QY 271 AlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu 284

DB 1464 GCCTATCTCAACTGGATCTACATGCTGTGAGAGGTGAGCTG 1505

RESULT 9

AR232520

Grimaldi, C.J., Hillan, K.J., Paoni, N.F., Roy, M.A. and Watanabe, C.K.
 Secreted and transmembrane polypeptides and nucleic acids encoding
 the same

Patent: WO 0078961-A 274 28-DEC-2000;

Genentech Inc. (US)

location/Qualifiers

1. .2063

/organism="Homo sapiens"

/mol_type="genomic DNA"

/db_xref="taxon:9606"

Qy 271 AlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu 284
Db 1478 GCCTATCTCACTGGGATCTACAATGTCGGAAGGCTGAGCTG 1519

RESULT 10
AX207967
LOCUS AX207967 2079 bp DNA linear PAT 31-AUG-2001
DEFINITION Sequence 71 from Patent WO0157194.
ACCESSION AX207967
VERSION AX207967.1 GI:15422563
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1
REFERENCE
AUTHORS Madison, E.L., Ong, E.O. and Yeh, J.C.
TITLE Nucleic acid molecules encoding transmembrane serine proteases, the
encoded proteins and methods based thereon
JOURNAL Patent: WO 0157194-A 71 09-AUG-2001;
CORVAS INTERNATIONAL, INC. (US)
FEATURES
source Location/Qualifiers
1..2079
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
251..1522
/notes="Nucleotide sequence encoding transmembrane
protease, serine 4 (TMPRSS4)"
/codon_start=1
/protein_id="CAC60389.1"
/db_xref="GI:15422564"
/translation="MSNPKANPNVSPWRPSESVCIPITIIALLSLASIIIVVVLIKVILD
KYIFLCQQLHFIIPKQLCGELDPLGDEHECVKSPFEGPAVAVRLSKORSTLOVL
DSATGNWGFACPDNTEALAEATACQMGYSKSPTRFAVIGPDQDLVVEITENSQEL
RMRNSGCPCLSGSLVSLACLAKSLKTPRVVGGEASVDSNPQVSIQVDKQHVCGG
IILDHFWLTAACHCRKHPTDVPNWKVRAGSKLGSPLAVAKIIIEBNPMYPKND
IALLKLOPLTPSGTVRPICLPFDEBELPATPLMILIGWFTKQNGKMSDILLQASV
OYLDSTRNADADAYQGEVTEKMKAGIPGGVDTCQDSSGGPLMYQSDQWVVVGVISW
GYGCGFSTPTGYTTCYSAYLNNIYNWKAEL"

BASE COUNT 489 a 594 c 575 g 421 t
ORIGIN

Alignment Scores:
Pred. No.: 1..16e-121 Length: 2079
Score: 1266.00 Matches: 228
Percent Similarity: 98.29% Conservative: 2
Best Local Similarity: 97.44% Mismatches: 4
Query Match: 79.12% Indels: 0
DB: 6 Gaps: 0

US-09-607-745-9 (1-292) x AX207967 (1-2079)

Qy 51 LysIleValGlyGlyTyrAlaLeuAspValAspSerTrpProTrpGlnValSerIleGln 70
Db 818 CQTGTGTGGTGGTGGGAGGAGGCCTCTGTGGATTCTTGGCCTTGGCAGGTGAGTCAC 877

Qy 71 TyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValIleuThrAla 90
Db 878 TAGCAACAACACACAGCTGTGTGGAGGGAGCATCTGGACCCCATCTGGGTCTCATCGGCA 937

Qy 91 AlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySerAsp 110
Db 938 GCCCACTGCTTCAGGAACAACATACCGATGTGTCACTGGAAGTGGGGCAGGCTCAGAC 997

Qy 111 LysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleIleGluPheAsnPro 130
Db 998 AAATGGGCAGCTTCCCATCCCTGGCTGTGGCCCAAGATCATCATCATTTGAATTCACCCC 1057

Qy 131 MetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPheSer 150
1058 ATGTACCCCAAGACAACATGATCCCTCATGAATCCCTCATGAATCCCTCATTTCTCA 1117

```

Y 151 GlyThrValArgProIleCysLeuProPhePheAspGluGluLeuThrProAlaThrPro 170
b 1118 GGCACAGTCAGCCCATCTGCTGCGCCCTCTTTGATGAGAGCTCACTCCAGCCACCCCA 1177
Y 171 LeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyGlyLysMetSerAspIleLeu 190
b 1178 CTCTGATCATTTGGATGGGCTTTACGACAGCAGATGGAGGAGATGCTGACATACG 1237
Y 191 LeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAlaThrGln 210
b 1238 CTGACAGGCGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAG 1297
Y 211 GlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThrCys 230
b 1298 GGGGAAGTCACCGAAGATGATGTGTGAGGATCCCGGAGGAGGAGGAGGAGGAGGAGGAG 1357
Y 231 GlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGlyIle 250
b 1358 CAGGCTGACAGTGGTGGGCTCTGATGATGATGATGATGATGATGATGATGATGATGATG 1417
Y 251 ValSerTrpGlyTrpGlyCysGlyGlyProSerThrProGlyValThrLysValSer 270
b 1418 GTTAGCTGGGCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1477
Y 271 AlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu 284
b 1478 GCCTATCTCACTGATCTACATGCTGTGGAAGGCTGAGCTG 1519

RESULT 11
P216312 AF216312 2079 bp mRNA linear PRI 07-FEB-2000
DEFINITION Homo sapiens type II membrane serine protease mRNA, complete cds.
ACCESSION AF216312
VERSION AF216312.1 GI:6911218
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
Smeekens,S.S., Lorimer,D.D., Wang,E., Hou,J. and Linnevers,C.
MT-SP2, a novel type II membrane serine protease expressed in
trachea, colon, and small intestine: identification, cloning, and
chromosomal localization
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2079)
AUTHORS Smeekens,S.S., Lorimer,D.D., Wang,E., Hou,J. and Linnevers,C.
TITLE Direct Submission
JOURNAL Submitted (14-DEC-1999) Axys Pharmaceuticals, Inc, 180 Kimball Way,
South San Francisco, CA 94080, USA
FEATURES
Location/Qualifiers
1..2079
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
251..1522
/note="MT-SP2"
/codon_start=1
/product="type II membrane serine protease"
/db_xref="GI:6911219"
/translation="MSNFCANVPSPRSESVGPIIILALISLAIIVVLKIVLD
KIYFLCGPLHPIPRKQCDLDELCPGEDEEHCVKFPFGFAVAVRLSKORSTLQVL
DSATGNWFSAEDNFTALATACRQMGYSKPTFAVEIGFDODLVETNSQEL
RMRNSGFLCLAGSLVSLHCLAGSLKTPRVVGGSEASVDSWPQVSTQYDKHVCQG
SLIDPHWLTAHCRKRTDVFNVKVRAGSKLGSFPLSAKILIIIEPNMYPKND
IALMKLPFLFSGTVRICLIPFDEELTPATPLMIGWFTKQNGKMSDILIQASV
QVIDTNCADADYQGVTEKMKCAGIPGGVDVTCQDGGSLMYSQDQWVGVTVSW
GYDGGCPSTPGVYTKVSAYLNIWVNWKAEL"
BASE COUNT 489 a 594 c 575 g 421 t
alignent Scores:

```

```

Pred. No.: 1.16e-121 Length: 2079
Score: 1266.00 Matches: 228
Percent Similarity: 98.29% Conservative: 2
Best Local Similarity: 97.44% Mismatches: 4
Query Match: 79.12% Indels: 0
DB: 9 Gaps: 0

US-09-607-745-9 (1-292) x AF216312 (1-2079)
QY 51 LysIleValGlyGlyTyrAlaLeuAspValAspSerTyrProTyrGlnValSerIleGln 70
Db 818 CGTGTGGGGTGGGAGGAGCCCTCTGTGATCTTTGGCCCTTGGACGTGAGCATCCAG 877
QY 71 TyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThrAla 90
Db 878 TACGACAAACAGCAGCTCTGTGGAGGAGCATCTCGACCCCTCCCTGGTCTCTCAGCGCA 937
QY 91 AlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySerAsp 110
Db 938 GCCCATGCTTTACGAAAACATACCGATGTGTTCATCGAAAGTCCGGGCGAGGCTCAGAC 997
QY 111 LysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleGluPheAsnPro 130
Db 998 AAATGGGAGCTTCCCATCTCTGGCTGTGGCAAGATCATCATTAATGAATCAACCCC 1057
QY 131 MetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPheSer 150
Db 1058 ATGTACCCCAAGACAATGACATCGCTCATGAAGCTGCAGTTCCTCCATCTCCTCTCA 1117
QY 151 GlyThrValArgProIleCysLeuProPhePheAspGluGluLeuThrProAlaThrPro 170
Db 1118 GGCACAGTCAGGCGCATCTGCTGCTCTCTTTGATGAGGAGCTCACTCCAGCCACCCCA 1177
QY 171 LeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyGlyLysMetSerAspIleLeu 190
Db 1178 CTCTGATCATTTGGATGGGCTTTACGAGCAGAAATGAGGAGGAGATGCTGTGACATACG 1237
QY 191 LeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAlaThrGln 210
Db 1238 CTGACAGGCTCAGTCCAGGTCATTGACAGCACACGCTGCAATGACAGCAGTCCGTACCG 1297
QY 211 GlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThrCys 230
Db 1298 GGGGAAGTCACCGAAGATGATGTGTGAGGATCCCGGAGGAGGAGGAGGAGGAGGAGGAG 1357
QY 231 GlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGlyIle 250
Db 1358 CAGGCTGACAGTGGTGGGCTCTGATGATGATGATGATGATGATGATGATGATGATGATG 1417
QY 251 ValSerTrpGlyTrpGlyCysGlyGlyProSerThrProGlyValThrLysValSer 270
Db 1418 GTTAGCTGGGCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1477
QY 271 AlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu 284
Db 1478 GCCTATCTCACTGATCTACATGCTGTGGAAGGCTGAGCTG 1519

RESULT 12
AF179224 AF179224 2081 bp mRNA linear PRI 08-JUN-2000
LOCUS Homo sapiens transmembrane serine protease 3 (TMPRSS3) mRNA,
complete cds.
DEFINITION
ACCESSION AF179224
VERSION AF179224.1 GI:8347148
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 2081)
Wallrapp,C., Hähnel,S., Muller-Pillasch,F., Burghardt,B.,
Iwamura,T., Ruthenburger,M., Lerch,M.M., Adler,G. and Gress,T.M.
A novel transmembrane serine protease (TMPRSS3) overexpressed in

```

pancreatic cancer
Cancer Res. 60 (10), 2602-2606 (2000)
MEDLINE
20283276
PUBMED
10825129

REFERENCE
2 (bases 1 to 2081)
Waller, C. and Gress, T.M.

AUTHORS
Direct Submission
TITLE
Submitted (20-AUG-1999) Internal Medicine I, University of Ulm,
JOURNAL
Robert-Koch-Street 8, Ulm, Baden-Wuerttemberg 89081, Germany

FEATURES
Location/Qualifiers

source
1..2081
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="11"
/map="11q23.3"
/tissue_type="pancreatic carcinoma"
1..2081
/gene="TPRSS3"
215..1528
/gene="TPRSS3"
/codon_start=1
/product="transmembrane serine protease 3"
/protein_id="AAF74526.1"
/db_xref="GI:8347149"
/translation="MLQDPSDPSLSDVPLKPRIPMETPRKVGIPPIIALLSIA
SIIIVVLKIVLDKFIYLCQQLHFIKQLCDGLDCLPGLDEDEHCKVSPFEGP
AVLSKDRSTLOVLDATNWSFDFNTEALAEATACQMGYSKPTFAVEIGDPQ
DLDVETITNSBELNRNNSGFCLSGLVSLHCLAGCKSLKTPRVVGGEEASVDSWPM
QVSIQYDKHVCVCGSILDPHWLTAACFKRHTDVFNWKRAGSDKLGSPFSLAVAKI
IIIFNPMVKNDKIALMKLOPELTFSGVTRPICLPFDEELTPATPLIINGPFTKO
NGKQSDILLQSVQVIDSTRNADDAVQGEVERKMCAGIPFGVDTCQGSGGGLPM
YQSDQHVVGIVSWGIGCGGSPGVTIKVSLYINWVKAEL"
215..307
/gene="TPRSS3"
/note="Region: cytoplasmic domain"
308..373
/gene="TPRSS3"
/note="Region: transmembrane domain"
374..1525
/gene="TPRSS3"
/note="Region: extracellular domain"
602..604
/gene="TPRSS3"
/note="glycosylation site"
746..748
/gene="TPRSS3"
/note="glycosylation site"
order(800..802,1142..1144)
/gene="TPRSS3"
/note="disulfide bond"
order(902..904,950..952)
/gene="TPRSS3"
/note="disulfide bond"
947..949
/gene="TPRSS3"
/note="active site"
1082..1084
/gene="TPRSS3"
/note="active site"
order(1280..1282,1328..1330)
/gene="TPRSS3"
/note="disulfide bond"
1373..1375
/gene="TPRSS3"
/note="active site"

BASE COUNT 484 a 597 c 576 g 424 t

REGION

Alignent Scores:
red. No.: 1.17e-121 Length: 2081
core: 1266.00 Matches: 228
percent Similarity: 98.29% Conservative: 2

Best Local Similarity: 97.44% Mismatches: 4
Query Match: 79.12% Indels: 0
DB: 9 Gaps: 0
US-09-607-745-9 (1-292) x AF179224 (1-2081)
QY 51 LysIleValGlyGlyTyrAlaLeuAspValAspSerTrpProTrpGlnValSerIleGln 70
Db 824 CGTGTGCTGGTGGGAGGAGGCTCTGTGGATTTCTTGGCTTGGCAGGTTCAGCATCCAG 883
QY 71 TyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThrAla 90
Db 884 TAGCACAACAGCACGCTCTGTGGAGGAGCATCTGGACCCACCTGGGTCTTCAGGGCA 943
QY 91 AlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySerAsp 110
Db 944 GCCCACTGCTTCAGGAACAATACCGATGTGTCACTGGAAGGTGGGCGGCTCAGAC 1003
QY 111 LysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleIleGluPheAsnPro 130
Db 1004 AAACCTGGGAGCTTCCCATCCCTGGCTGTGGCAAGATCATCATATTGAATCAACCC 1063
QY 131 MetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPheSer 150
Db 1064 ATGATCCCAAGACAATACATCGCCCTCATGAAGCTGCAGTCCCATCTCTTCTCA 1123
QY 151 GlyThrValArgProIleCysLeuProPhePheAspGluGluLeuThrProAlaThrPro 170
Db 1124 GGCACAGTCAGGCCCATCTGTCTGCCCTTCTTGTATGAGAGCTCACTCCAGCACCCCA 1183
QY 171 LeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyGlyLysMetSerAspIleLeu 190
Db 1184 CTCGTGATCATTTGGATGGGGCTTTACGAAGCAATGGAGGAAGATGCTGCATACTG 1243
QY 191 LeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAspAlaTyrGln 210
Db 1244 CTGCGAGGTCTAGTCTCAGGTCAATTACAGCACACGCTGCAATGCAGACATGCTGACCAG 1303
QY 211 GlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThrCys 230
Db 1304 GGGGAAGTCACCGAGAAGATGATGTGCAGCATCCCGAAGGGGGTGTGGACACTGC 1363
QY 231 GlnGlyAspSerGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGlyIle 250
Db 1364 CAGGGTGACGTGTGGGCCCTCTGATACCAATCTGACAGTGGCATGTGTGGGCATC 1423
QY 251 ValSerTrpGlyTyrGlyCysGlyProSerThrProGlyValTyrThrLysValSer 270
Db 1424 GTTAGCTGGGCTATGGCTGCGGGGGCCCGAGACCCCGAGGAGTATACACCAAGGTCTCA 1483
QY 271 AlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu 284
Db 1484 GCCTATCTCACTGGATCTACAAATGCTCTGGAAGGTGAGCTG 1525

RESULT 13

BC011703

LOCUS

DEFINITION

Homo sapiens, similar to mosaic serine protease, clone MGC:19490

IMAGE:3610695, mRNA, complete cds.

ACCESSION

BC011703

VERSION

BC011703.1 GI:5079794

KEYWORDS

MGC.

SOURCE

Homo sapiens (human)

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 2088)

AUTHORS

Strausberg, R.

TITLE

Direct Submission

JOURNAL

Submitted (30-JUL-2001) National Institutes of Health, Mammalian

Gene Collection (MGC), Cancer Genomics Office, National Cancer

Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,

USA

REMARK
COMMENT

NIH-MGC Project URL: <http://mgc.nci.nih.gov>
 Contact: MGC help desk
 Email: cgabbs-r@mail.nih.gov
 Tissue Procurement: ATCC
 cDNA Library Preparation: Rubin Laboratory
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: National Institutes of Health Intramural
 Sequencing Center (NISC),
 Gaithersburg, Maryland;
 Web site: <http://www.nisc.nih.gov/>
 Contact: nisc.mgc@nih.gov
 Shevchenko, Y., Wetherby, K.D., Beckstrom-Sternberg, S.M.,
 Benjamin, B., Blakesley, R.W., Bouffard, G.G., Brinkley, C., Brooks, S.,
 Dietrich, N.L., Guan, X., Gupta, J., Ho, S.-L., Karlins, E., Legaspi, R.,
 Lim, M., Maduro, Q.L., Masiello, C., Mastrian, S.D., McCloskey, J.C.,
 McDowell, J., Pearson, R., Snyder, B., Stantripop, S., Thomas, P.J.,
 Tongson, E.E., Touchman, J.W., Tsurgeon, C., Vogt, J.L., Walker, M.A.,
 Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/ILNL at: <http://image.llnl.gov>
 Series: IBAL Plate: 27 Row: f Column: 19
 This clone was selected for full length sequencing because it
 passed the following selection criteria: matched mRNA gi: 8347148.

FEATURES
source

1..2088
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="MGC:19490 IMAGE:3610695"
 /tissue_type="Pancreas, adenocarcinoma"
 /clone_lib="NIH_MGC_39"
 /lab_host="DH10B-R"
 /note="Vector: pOTE7"
 217..1530
 /codon_start=1
 /product="Similar to mosaic serine protease"
 /protein_id="AAH11703.1"
 /db_xref="gi:15079795"
 /translation="MLQPDSDQPLNSLDVPLRPKIPMETFRKVGIPILIALLSLA
 SIIVLVILKVLQYFLQCPQLHFIKQKLCDELCPGEBDEHCVKFPFPGPAV
 AVRLSKSDTQLDLSANGNFCACDFNTEALACQMSKSPKFFRAVEIGPDQ
 DLDVETIENSQELMRNSGCLSGSLVSLHCLACSKLTPRVVGERASVDSWPW
 QVSIQKHQCVGGSLDHPVLTAACFRKHTDVFNNKVRAGSDKLGSPSLAVAKI
 IIFENPMYKNDIALMKLQPLTFSTVTRPICLPFDDELTPATPLWIIIGWGTQK
 NGKMSDILLOASVQVIDSTRCNADDAVQGEVTEKMCAGIPGGVDTCQDSGGLM
 YQSDQWVVVGIWVGCGGSPTRGVYTKVSAYLNIWVWKAEI"

CDS

RESULT 14
 E13203
 LOCUS Human cDNA encoding a serine protease. 699 bp DNA linear PAT 27-APR-1998
 DEFINITION Human cDNA encoding a serine protease.
 ACCESSION E13203
 VERSION E13203.1 GI:3252008
 KEYWORDS JP 1997149790-A/4.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 699)
 Tsuruoka, N., Yamashiro, K., Tsujimoto, M. and Yamaguchi, M.
 NEW SERINE PROTEASE
 Patent: JP 1997149790-A 4 10-JUN-1997;
 SUNTORY LTD
 OS Homo sapiens (human)
 PN JP 1997149790-A/4
 PD 10-JUN-1997
 PF 24-JUL-1996 JP 1996212196
 PR 29-SEP-1995 JP 95P 275105
 PI TSURUOKA NOBUO, YAMASHIRO KYOKO, TSUJIMOTO MASAFUMI, PI
 YAMAGUCHI MARE
 PC C12N15/09, C07H21/04, C07K14/47, C12N1/21, C12N5/10, PC
 C12N9/52//A61K38/46,
 PC (C12N15/09, C12R1:91), (C12N1/21, C12R1:19), (C12N9/52, C12R1:19),
 PC (C12N9/52,
 CC CC
 CC strandedness: Double;
 CC topology: Linear;
 FH Key Location/Qualifiers
 FT source 1..699
 FT /organism="Homo sapiens"
 FT /cell_type="colon cancer-derived cell" FT
 FT /cell_line="COLO 201"
 FT /clone="SP60"
 FT mat_peptide 1..699

1006 AAACCTGGGAGCTTCCCATCCTGGCTGTGGCAAGATCATCATGATTAATCAACCCC 1065
 131 MetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPheSer 150
 1066 ATGTACCCCAAGACAATGACATCGCTCATGAGTGCAGTTCCTCCACTCATTCTCTCA 1125
 151 GlyThrValArgProIleCysLeuProPhePheAspGluGluLeuThrProAlaThrPro 170
 1126 GGCACAGTCAGCCCACTGTCTGCCCTCTTTGATGAGGAGCTCATTCCAGCCACCCCA 1185
 171 LeuTrpIleIleGlyTyrGlyPheThrLysGlnGlyLysMetSerAspIleLeu 190
 1186 CTCTGGATCATTTGGATGGGCTTTACGAAGCAGAATGAGGAGGAGATGCTTGACATCTG 1245
 191 LeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAlaThrGln 210
 1246 CTGAGGGCGTCAGTCCAGGTCAATGACAGCACGCGTGCATGACCAATCTGACCAATGCGTAC 1305
 211 GlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThrCys 230
 1306 GGGGAAGTCCACGAGAGATGATGTGTGACAGCATCCCGGAGGGGTGTGGACACCTGC 1365
 231 GlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValGlyIle 250
 1366 CAGGTCACAGTGGTGGGCGCTGATGTACCAATCTGACCAATGCGTACGATGCGTGGGCATC 1425
 251 ValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyValTyrThrIleValSer 270
 1426 GTTAGCTGGGCTATGGCTGGGGGCGGCGGACCCAGGAGTATACCAAGGTCTCA 1485
 271 AlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu 284
 1486 GCCTATCTCACTGAGTCAATGATGTCTGGAGGCTGAGCTG 1527

ALIGNMENT Scores:
 red. No.: 1.17e-121 Length: 2088
 core: 1266.00 Matches: 228
 Percent Similarity: 98.29% Conservative: 2
 best Local Similarity: 97.44% Mismatches: 4
 Query Match: 79.12% Indels: 0
 Gaps: 9

IS-09-607-745-9 (1-292) x BC011703 (1-2088)
 Y 51 LysIleValGlyGlyTyrAlaLeuAspValAspSerTrpProTrpGlnValSerIleGln 70
 b 826 CGTGTGGTGGTGGGAGGAGGAGCCCTCTGTGGATTCTTGCCCTTGGCAGGTGAGCATCCAG 885
 Y 71 TyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThrAla 90
 b 886 TAGCAACAACGACAGCTTGTGGAGGAGGAGCATCTTGGACCCCACTGGGTCTCAGGCA 945
 Y 91 AlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySerAsp 110
 b 946 GCCCACTGCTTCAGGAACATACCATGCTGTCTCACTGAGAGGTGGCGGAGGCTCAGAC 1005
 Y 111 LysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleGluPheAsnPro 130

```
FEATURES             Location/Qualifiers
     source                1..699
     /product="Serine protease".
     /organism="Homo sapiens"
     /mol_type="genomic DNA"
     /db_xref="taxon:9606"
BASE COUNT      157 a 190 c 209 g 143 t
ORIGIN
1  Homo sapiens (human)
2  Homo sapiens
3  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
4  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
5
6  Suendermann, B., Hofmann, U., Matzku, S. and Wilbert, O.
7  Seripancrin
8  Patent: WO 0104141-A 3 18-JAN-2001;
9  MERCK PATENT GmbH (DE)
10
11  FEATURES
12  source
13  Location/Qualifiers
14  1..1479
15  /organism="Homo sapiens"
16  /mol_type="genomic DNA"
17  /db_xref="taxon:9606"
18  1..1479
19  /note="unnamed protein product"
20  /codon_start=1
21  /protein_id="CAC28480.1"
22  /db_xref="GI:12710818"
23  /translation="MDPDSQPLNSLDVPLKRPIMETERKVGIPITIIALLSLASI
24  IIVVLIKVLKYFLCGQPLHFTPRKLCGDGELDCPLGEDEHCHVKSFFEGPAVAV
25  RLSKDRSTLOVLDSATGWFSAFCDFNTEALAEACRQMGYSKPTFEAVIGPDOL
26  DVAITENSQBLRMNSGFCLSGLSLHCLACGKSLKTRVVGGEASVDSWVQV
27  SIQDKQHVCGSILDPHWLTAACFRKTDVFNWKRAGSDKLGSPFLAVAKILII
28  IEFNPTPKNDIALMKLQPELTFTGTVPICLPFFDBELTPATLMTIGRGTKQNG
29  GKMSDILLOASVQVIDSTRCNADDAVQGEVTEKMKACIGPEGVDTCQDGGGPMYQ
30  SDQHWVIGVSWGVCSPSTPGVYTKVSAVLNWIYNWKDRTIQRSCNSPTGLVIO
31  QPAPVLMELRSARHTRILTCVCFSLSPALQPLSLCPFPWD"
32
33  BASE COUNT      329 a 425 c 407 g 318 t
34  ORIGIN
35
36  Alignment Scores:
37  Pred. No.:      1,7e-120      Length:      1479
38  Score:          1253.00      Matches:      225
39  Percent Similarity: 98.27%      Conservative: 2
40  Best Local Similarity: 97.40%      Mismatches: 4
41  Query Match:      78.31%      Indels: 0
42  DB:              6              Gaps: 0
43
44  US-09-607-745-9 (1-292) x AX076192 (1-1479)
45
46  QY 51 LysleValGlyGlyTyrAlaLeuAspValAspSerTrpProTrpGlnValSerIleGln 70
47  DB 604 CGTGTGTGGTGGGAGGAGGCGCTCTGTGATTTCTTGGCTTGGCAGGTGAGATCCAG 663
48
49  QY 71 TyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThrAla 90
50  DB 664 TACGACAAACAGCACGCTCTGTGGAGGGAGCATCTTGGACCCCTGGGTCTCAGGCA 723
51
52  QY 91 AlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySerAsp 110
53  DB 724 GCCCAGTCTTCAGAAACATACCGATGTGTTCACCTGGAAGGTGCGGCGAGGTTCAGAC 783
54
55  QY 111 LysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleGluPheAsnPro 130
56  DB 784 AAACCTGGGAGCTTCCCATCCCTGGCTGTGGCCCAAGATCATCATCAATGAATTCACCCC 843
57
58  QY 131 MetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPheSer 150
59  DB 844 ATGTACCCCAAGACAAATGACATCGCCCTCATGAAGCTGCAGTTCACCATCTCTTCTCA 903
60
61  QY 151 GlyThrValArgProIleCysLeuProPheAspGluGluLeuThrProAlaThrPro 170
62  DB 904 GGCACATCTAGGCCCATCTGTCTGCCCTTCTTGTATGAGGAGCTCACTCAGCCACCCCA 963
63
64  QY 171 LeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyGlyLysMetSerAspIleLeu 190
65  DB 964 CTCTGGATCATTTGGATGGGCTTTACGACAGAGATCGAGGAGAGATGCTCTGACATCTG 1023
66
67  QY 191 LeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAspAlaTyrGln 210
68  DB 1024 CTGCAGGCGTCAGTCCAGGTCAATTGACAGCACACGGTGCAATGCAGACGATGCGTACCAG 1083
```

```
y 211 GlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThrCys 230
b 1084 GGGGAAGTCACCGAGAAGATGATGTGTCAGGCATCCCGGAAGGGGTGTGGACACCTGC 1143
y 231 GlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGlyIle 250
b 1144 CAGGCTGACAGTGTGGGCCCTGATGTACCAATCTGACCCAGTGGCATGTGTGGGCATC 1203
y 251 ValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyValTyrThrIysValSer 270
b 1204 GTTAGCTGGGGCTATGGCTGCGGGGCCCGAGACCCCGAGAGTATACACCAAGGTCTCA 1263
y 271 AlaTyrLeuAsnTrpIleTyrAsnValTrpLys 281
b 1264 GCCTATCTCAACTGGATCTACAAATGCTGGAAG 1296
```

earch completed: November 23, 2003, 10:40:15
Job time : 3705.55 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

M protein - nucleic search, using frame_plus_p2n model

un on: November 23, 2003, 07:49:55 ; Search time 263.081 Seconds
(without alignments)
2996.170 Million cell updates/sec

itle: US-09-607-745-9

erfect score: 1600

equance: 1 MDSGSGSQSKRLLLLVVSN.....LNWYNVWKAELSRHHHHH 292

coring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

earched: 2552756 seqs, 1349719017 residues

otal number of hits satisfying chosen parameters: 5105512

inimum DB seq length: 0

aximum DB seq length: 2000000000

ost-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

ommand line parameters:

MODEL=frame+p2n.model -DEV=xlh
Q=/cgn2.1/USPTO_spool/US09607745/runat 21112003 144344 22176/app query.fasta_1.1038
DB=N Geneseq 19Jun03 -QFMT=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOFCI=0
LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -WATRX=blotsum62 -TRANS=human40.cdi
LIST=45 -DOCALIGN=200 -THR SCORE=pot -THR MAX=100 -THR MIN=0 -ALIGN=15
MODE=LOCAL -OUTFMT=pro -NORM=ext -HEAPSIE=500 -MINLEN=0 -MAXLEN=2000000000
USER=US09607745 @CGN 1.1 401 @runat 21112003 144344 22176 -NCFU=6 -ICPU=3
NO MAP -LARGQUERY -NEG SCORE=0 -WAIT -DSPBLOCK=100 -LONGLOG
DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
FGAEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

atabase : N_Geneseq 19Jun03.*

1: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*
25: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed.

and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1593	99.6	1189	24	AAI64290	Protease D-G catay
2	1266	79.1	1305	22	AA02556	Human seripancrin
3	1266	79.1	1314	25	AAH76354	Lung cancer-associ
4	1266	79.1	1854	22	AAH99574	Human protein enco
5	1266	79.1	2038	20	AAH87154	Human protease HUP
6	1266	79.1	2063	21	AAH37099	Human PRO1570 (UNQ
7	1266	79.1	2063	22	AAH46089	Human DNA encoding
8	1266	79.1	2063	22	AAH92113	Human PRO1570 cDNA
9	1266	79.1	2063	22	AAH43396	DNA encoding prote
10	1266	79.1	2063	24	ABH74433	Human cDNA encodin
11	1266	79.1	2063	24	ABK11090	cDNA encoding tumo
12	1266	79.1	2063	25	ACA57847	Human PRO1570 cDNA
13	1266	79.1	2063	25	ACA58865	cDNA encoding huma
14	1266	79.1	2063	25	ACA60418	Novel human secret
15	1266	79.1	2063	25	ACA63428	cDNA encoding huma
16	1266	79.1	2063	25	ABX98317	Human cDNA encodin
17	1266	79.1	2063	25	ABX98819	Novel human secret
18	1266	79.1	2063	25	ACA58864	Human secreted/tra
19	1266	79.1	2063	25	ABX97908	Human PRO polynucl
20	1266	79.1	2063	25	ABX78692	Human PRO polynucl
21	1266	79.1	2063	25	ABX75705	Human cDNA encodin
22	1266	79.1	2063	25	ABX76910	Human PRO polynucl
23	1266	79.1	2063	25	ABX16750	Human cDNA encodin
24	1266	79.1	2079	22	AAH13169	Human transmembran
25	1266	79.1	2079	24	AAI72975	CUA8 cDNA. Homo s
26	1266	79.1	2079	25	ABH57763	cDNA encoding huma
27	1266	79.1	2081	24	AAI72976	CUA8 preferred cDN
28	1266	79.1	2165	24	ABZ35523	Human gene express
29	1266	79.1	2307	24	ABH76529	cDNA encoding huma
30	1264	79.0	699	18	AAH79127	Human serine prote
31	1259	78.7	2121	24	AAI84284	Human serine prote
32	1253	78.3	1479	22	AAH02557	Human seripancrin
33	1253	78.3	2070	21	AAZ90471	Cancer specific ge
34	1248	78.0	2137	22	AAH13114	Human membrane-typ
35	1246	77.9	1795	22	AAH26880	Human cDNA encodin
36	1230	76.9	1281	21	AAH37361	Human colorectal c
37	918	57.4	1084	22	AAH34953	Human colon cancer
38	806	50.4	1008	20	AAH04381	Human secreted pro
39	765	47.8	1130	22	AAH01366	Recombinant human
40	754.5	47.2	1130	22	AAH77000	Fusion gene of pro
41	746	46.6	1169	21	AAH87795	Activation constru
42	746	46.6	1169	22	AAH55267	Nucleotide sequenc
43	740.5	46.3	1166	22	AAH02991	Zymogen activation
44	672.5	42.0	1052	21	AAH87798	Activation constru
45	672.5	42.0	1052	22	AAH55270	Nucleotide sequenc

ALIGNMENTS

RESULT 1

AAI64290

ID AAI64290 standard; DNA; 1189 BP.

XX AAI64290;

AC AAI64290;

XX 07-MAY-2002 (first entry)

XX Protease D-G cataytic domain fusion gene construct encoding sequence.

DE Serine protease; D-G; human; zymogen; enzyme; cytostatic;

XX antiinflammatory; dermatological; anticoagulation; cancer;

KW skin disorder; neuropathic pain; inflammatory disorder;

KW coagulation diathesis; thrombosis; laundry detergent; skin care;

XX gene therapy; gene; ds.

XX Homo sapiens.

C antiparkinsonian; and immunostimulant. The proteins and polynucleotides
C encoding them can be used in gene therapy, antisense therapy and vaccine
C production. The proteins and polynucleotides are useful for screening for
C agonists or antagonists of a protein and for the treatment and diagnosis
C of disorders associated with the activity of a protein e.g. inflammation,
C rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,
C neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal
C infections, autoimmunity, genetic diseases, haematopoietic disorders,
C anaemia, platelet disorders, thrombocytopaenia, wounds, burns, ulcers,
C osteoporosis, severe combined immunodeficiency, eczema, allergic
C rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression,
C Alzheimer's disease, Parkinson's disease, neurodegenerative and
C neurological disorders.

Q Sequence 1854 BP; 397 A; 513 C; 495 G; 449 T; 0 other;

Alignment Scores:
red. No.: 2,31e-120 Length: 1854
core: 1266.00 Matches: 228
Percent Similarity: 98.29% Conservative: 2
Best Local Similarity: 97.44% Mismatches: 4
Query Match: 79.12% Indels: 0
B: 22 Gaps: 0

S-09-607-745-9 (1-292) x AAH99574 (1-1854)

Y 51 LysIleValGlyGlyTyrAlaLeuAspValAspSerTrpTrpGlnValserIleGln 70
b 1605 CGTGTGGTGGTGGGAGGAGCCCTCTGTGGATTCTTGGCCCTGGCAGTCAGCATCCAG 1546

Y 71 TyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThrAla 90
b 1545 TACGCAACACACACCTCTGTGGAGGAGCATCTGGACCCCTCTGGTCTCTACGGCA 1486

Y 91 AlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySerAsp 110
b 1485 GCCCACTCTTCAGAAACATACCATGTTCACTGGAAGTGGGGCAGGCTCAGAC 1426

Y 111 LysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleIleGluPheAsnPro 130
b 1425 AAACCTGGGCGAGTCTCCCATCCCTGGCTGTGGCAAGATCATCATTAATCAACCCC 1366

Y 131 MetTyrProLysAspAspIleAlaLeuMetLysLeuGlnPheProLeuThrPheSer 150
b 1365 ATGTACCCCAAGCAATGACATCCCTCTCACTGAAGTGGGGCAGGCTCAGAC 1306

Y 151 GlyThrValArgProIleCysLeuProPhePheAspGluGluLeuThrProAlaThrPro 170
b 1305 GGCACAGTCAGGCCCATCTGTCTGCCCTCTTTGATGAGGAGTCACTCCAGCCACCCCA 1246

Y 171 LeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyGlyLysMetSerAspIleLeu 190
b 1245 CTCTGGATCATGTGGATGGGCTTACGAGCAGACACCGTGCATGACAGATGGGAGATCTCTGACATCTG 1186

Y 191 LeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAlaThrGln 210
b 1185 CTGAGGGGTGAGTCCAGTCCAGTTCATTCACAGCAGACCGTGCATGACAGATGGGAGATCTCTGACATCTG 1126

Y 211 GlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThrCys 230
b 1125 GGGGAAGTCACCGAAGATGATGTGTGAGGATCCCGAAGGGGGTGTGGACACTTGC 1066

Y 231 GlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGlyIle 250
b 1065 CAGGGTGACAGTGGTGGGCCCTGTATGTACCAATCTGACCACTGGCATGTGTGGGCAATC 1006

Y 251 ValSerTrpGlyTyrGlyCysGlyProSerThrProGlyValThrLysValSer 270
b 1005 GTTAGCTGGGGCTATGGTGTGGGGGGCCGAGCACCCCGAGGATATACACCAAGGTCTCA 946

Y 271 AlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu 284
b 945 GCCTATCTCACTGGATTTACATGTTCTGGAGGGTGTGAGCTG 904

RESULT 5
AAH87154
ID AAX87154 standard; cDNA; 2038 BP.
XX AC AAX87154;
XX DT 27-SEP-1999 (first entry)
XX DE Human protease HUPM-6 cDNA.
XX KW Serine protease; human; HUPM-6; cell proliferation; cancer;
XX OS immune disorder; inflammation; therapy; ss.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
XX FT CDS 200..1507
XX FT /*tag= a
XX PN WO9936550-A2.
XX PD 22-JUL-1999.
XX PF 12-JAN-1999; 99WO-US00655.
XX PR 16-JAN-1998; 98US-0008271.
XX PA (INCY-) INCYTE PHARM INC.
XX PI Bandman O, Corley NC, Guegler KJ, Hillman JL, Shah P;
XX PI Tang YF, Yue H;
XX DR WPI; 1999-430616/36.
XX DR P-PSDB; AAY06437.
XX PT Novel human protease molecules useful in the treatment of
XX PT developmental disorders and/or cancers
XX PS Claim 8; Page 86-87; 90pp; English.
XX CC This nucleotide sequence codes for HUPM-6 (see AAY06437), a novel
XX CC human protease. HUPM-6 cDNA was initially identified in Incyte
XX CC Clone 1337018 from the colon cDNA library COLANNOT13 using a
XX CC computer search for amino acid sequence alignments. The present
XX CC sequence is a consensus sequence derived from overlapping and/or
XX CC extended nucleic acid sequences: Incyte Clones 1271725 (TESTTUT02),
XX CC 1337018, 586982 and 58598 (UTRSNOT01). A fragment comprising
XX CC nucleotides 900-949 of the present sequence can be used for
XX CC hybridisation. This sequence encompasses an active site residue.
XX CC Northern analysis shows expression of HUPM-6 in gastrointestinal,
XX CC and male and female reproductive cDNA libraries. Approximately 65%
XX CC of these libraries are associated with neoplastic disorders, and
XX CC 22% with the immune response. The invention provides 12 new human
XX CC proteases, i.e. HUPM-1 to -12 (see AAY06432-43), and the
XX CC polynucleotides encoding them (see AAX87149-60). Also provided are
XX CC vectors, host cells and methods for producing HUPM polypeptides, as
XX CC well as agonists and antagonists of HUPM. Methods for treating or
XX CC preventing cell proliferative disorders and immune disorders using
XX CC HUPM or HUPM antagonists are claimed.

SQ Sequence 2038 BP; 462 A; 591 C; 569 G; 416 T; 0 other;

Alignment Scores:
Pred. No.: 2,61e-120 Length: 2038
Score: 1266.00 Matches: 228
Percent Similarity: 98.29% Conservative: 2
Best Local Similarity: 97.44% Mismatches: 4
Query Match: 79.12% Indels: 0
DB: 20 Gaps: 0

US-09-607-745-9 (1-292) x AAX87154 (1-2038)

R 20-OCT-1998; 98US-0104987.
 R 20-OCT-1998; 98US-0105000.
 R 20-OCT-1998; 98US-0105002.
 R 21-OCT-1998; 98US-0105104.
 R 22-OCT-1998; 98US-0105169.
 R 22-OCT-1998; 98US-0105266.
 R 26-OCT-1998; 98US-0105266.
 R 26-OCT-1998; 98US-0105693.
 R 26-OCT-1998; 98US-0105694.
 R 26-OCT-1998; 98US-0105807.
 R 27-OCT-1998; 98US-0105881.
 R 27-OCT-1998; 98US-0105882.
 R 27-OCT-1998; 98US-0106062.
 R 28-OCT-1998; 98US-0106023.
 R 28-OCT-1998; 98US-0106029.
 R 28-OCT-1998; 98US-0106030.
 R 28-OCT-1998; 98US-0106032.
 R 28-OCT-1998; 98US-0106033.
 R 28-OCT-1998; 98US-0106178.
 R 29-OCT-1998; 98US-0106248.
 R 29-OCT-1998; 98US-0106384.
 R 29-OCT-1998; 98US-0108500.
 R 30-OCT-1998; 98US-0106464.
 R 03-NOV-1998; 98US-0106856.
 R 03-NOV-1998; 98US-0108902.
 R 03-NOV-1998; 98US-0108905.
 R 03-NOV-1998; 98US-0108919.
 R 03-NOV-1998; 98US-0108932.
 R 03-NOV-1998; 98US-0108934.
 R 10-NOV-1998; 98US-0107783.
 R 17-NOV-1998; 98US-0108775.
 R 17-NOV-1998; 98US-0108779.
 R 17-NOV-1998; 98US-0108787.
 R 17-NOV-1998; 98US-0108788.
 R 17-NOV-1998; 98US-0108801.
 R 17-NOV-1998; 98US-0108802.
 R 17-NOV-1998; 98US-0108806.
 R 17-NOV-1998; 98US-0108807.
 R 17-NOV-1998; 98US-0108867.
 R 17-NOV-1998; 98US-0108925.
 R 18-NOV-1998; 98US-0108848.
 R 18-NOV-1998; 98US-0108849.
 R 18-NOV-1998; 98US-0108850.
 R 18-NOV-1998; 98US-0108851.
 R 18-NOV-1998; 98US-0108852.
 R 18-NOV-1998; 98US-0108858.
 R 18-NOV-1998; 98US-0108904.

(GETH) GENENTECH INC.

Baker K, Goddard A, Gurney AL, Smith V, Watanabe CK, Wood WI;

WPI; 2000-237871/20.

P-PSDB; AAY99417.

T New mammalian DNA sequences encoding transmembrane, receptor or
 T secreted PRO polypeptides, useful for screening of potential peptide or
 T small molecule inhibitors of the relevant receptor/ligand interactions

X Claim 2; Fig 155; 773pp; English.

X AAA37022 to AAA37144 encode the new isolated human transmembrane,
 X receptor or secreted PRO polypeptides given in AAY99340 to AAY99462. The
 X transmembrane and receptor PRO proteins can be used for screening of
 X potential peptide or small molecule inhibitors of the relevant
 X receptor/ligand interactions. The polypeptides and nucleotide sequences
 X encoding then have various industrial applications, including uses as
 X pharmaceutical and diagnostic agents. AAA37145 to AAA37330 represent
 X PCR primers and hybridization probes used in the isolation of the PRO
 X polypeptides from the present invention.

Q Sequence 2063 BP; 477 A; 591 C; 576 G; 419 T; 0 other;

Alignment Scores:

Pred. No.:	2.65e-120	Length:	2063
Score:	1266.00	Matches:	228
Percent Similarity:	98.29%	Conservative:	2
Best Local Similarity:	97.44%	Mismatches:	4
Query Match:	79.12%	Indels:	0
DB:	21	Gaps:	0

US-09-607-745-9 (1-292) x AAA37099 (1-2063)

QY	51	LysIleValGlyGlyTyrAlaLeuAspValAspSerTrpProTrpGlnValSerIleGln	70
DB	804	CTGTGTGGTGGGGAGAGGCGCTCTGTGGATTCTTGGCCCTGGCAGTCAGCATCCAG	863
QY	71	TyrAspLysGlnHisValCysGlySerIleLeuAspProHisTrpValLeuThrAla	90
DB	864	TACGACAAACAGCAGCTCTGTGGAGGAGCATCTTGGACCCCATCTGGGTCTCCAGGCA	923
QY	91	AlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySerAsp	110
DB	924	GCCCACTGCTTCAGAAACATACCGATGTGTCACTGGAAGTGGGGCAGGCTCAGAC	983
QY	111	LysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleIleIleIleIleIleIle	130
DB	984	AACTGGGCAGCTTCCCATCTGCTGGCTGTGGCAAGATCATCATTAATTCACCC	1043
QY	131	MetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPheSer	150
DB	1044	ATGTACCCCAAGACAATGACATCGCCTCATGAAGCTGCAGTTCCTCACTCTTCTCA	1103
QY	151	GlyThrValArgProIleCysLeuProPhePheAspGluLeuLeuThrProAlaThrPro	170
DB	1104	GGCACAGTCAGGCCCATCTGTCTGCCCTTCTTGTATGAGGAGCTCACTCCAGCCACCCA	1163
QY	171	LeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyGlyLysMetSerAspIleLeu	190
DB	1164	CTGTGATCATTTGATGGGCTTTACGAAGCAGATGAGGAAGATGTCTGACATCTG	1223
QY	191	LeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAlaTrpGln	210
DB	1224	CTGCAGGCGTCAGTCCAGTTCATTGACAGCACACGGTGCATGCAGACGATCGTACCAG	1283
QY	211	GlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThrCys	230
DB	1284	GGGGAAGTCACGAGAGATGATGTGTGAGGATCCCGAAGGGGTGTGACACCTGTC	1343
QY	231	GlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGlyIle	250
DB	1344	CAGGTGACAGTGTGGGCCCTCTGATGTACCAATCTGACCACTGGCATGTGGGCACTC	1403
QY	251	ValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyValTrpThrLysValSer	270
DB	1404	GTATAGTGGGCTATGGCTGGGGGGCCCGAGCACCCCGAGAGTATACACCAAGTCTCA	1463
QY	271	AlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu	284
DB	1464	GCATATCTCAACTGATCTACAAATGTCTGTGAAGGCTGAGCTG	1505

RESULT 7

AA546089

ID AAS46089 standard; cDNA; 2063 BP.

XX

AC AAS46089;

XX

DT 18-DEC-2001 (first entry)

XX

DE Human DNA encoding PRO polypeptide sequence #165.

XX

PRO polypeptide; mammal; tumour; cancer; human; cattle; horse; sheep; ss;
 dog; cat; pig; goat; rabbit; tumour necrosis factor alpha; TNF-alpha;
 blood; chondrocyte cell; cell proliferation; cell differentiation; colon;
 adrenal; lung; breast; prostate; rectum; cervix; liver; genetic disorder;
 PCR primer.

XX

I Pan J, Paoni NF, Roy MA, Smith V, Stewart TA, Tumas D;
I Watanabe CK, Williams PM, Wood WI;
X WPI; 2001-071395/08. R

Secreted and transmembrane proteins and nucleic acids designated PRO, useful as hybridization probes, in chromosome and gene mapping and gene therapy -

S Claim 2; Fig 151; 787pp; English.

C The present invention relates to secreted and transmembrane proteins.
C These proteins and the DNA encoding them may be used as hybridization
C probes, in chromosome and gene mapping and in the generation of
C anti-sense RNA and DNA. They may also be used to generate either
C transgenic animals or knockout animals which are in turn useful for
C development and screening of therapeutically useful reagents.
C The nucleic acids may also be used in gene therapy.

Q Sequence 2063 BP; 477 A; 591 C; 576 G; 419 T; 0 other;

Lignment Scores:

red. No.:	2.65e-120	Length:	2063
core:	1266.00	Matches:	228
Percent Similarity:	98.29%	Conservative:	2
est Local Similarity:	97.44%	Mismatches:	4
very Match:	79.12%	Indels:	0
DB:	22	Gaps:	0

'S-09-607-745-9 (1-292) x AAF54396 (1-2063)

51	LYS	ILE	VAL	GLY	GLY	TYR	ALA	LEU	ASP	VAL	ASP	SER	TRP	PRO	TRP	GLN	VAL	SER	ILE	GLN	70
804	CGT	GGT	TGG	TGG	GGG	GAG	GAG	GCG	CTC	TGT	GCG	AAT	TCT	TGG	CCCT	TGG	CAG	GT	CAG	CA	863
71	TYR	ASP	LYS	GLN	HIS	VAL	CYS	GLY	GLY	SER	ILE	LEU	ASP	PRO	HIS	TRP	VAL	LEU	THR	ALA	90
864	TAC	GAC	AAA	CAG	CAC	GCT	CTG	TGT	GAG	GAG	GAC	TCT	TGG	ACC	CCC	CACT	TGG	GTC	CTC	AC	923
91	ALA	HIS	CYS	PHE	ARG	LYS	HIS	THR	ASP	VAL	PHE	ASN	TRP	LYS	VAL	ARG	ALA	GLY	SER	ASP	110
924	GCC	CACT	GCT	PTC	AGG	AAA	CAT	CCG	GAT	GTT	TCA	ACT	TGG	AAG	TGC	GGG	CAG	GCT	CG	AC	983
111	LYS	LEU	GLY	SER	PHE	PRO	SER	LEU	ALA	VAL	ALA	LYS	ILE	ILE	ILE	LEU	PHE	ASN	PRO	130	
984	AAA	CT	TGG	CAG	CTT	CCC	AT	CCCT	TGG	CTG	TGG	CAC	AGAT	CAT	CAT	CAT	TGA	ATT	CA	CCC	1043
131	MET	TYR	PRO	LYS	ASP	ASN	ASP	ILE	ALA	LEU	MET	LYS	LEU	GLN	PHE	PRO	LEU	THR	PHE	SER	150
1044	ATG	TAC	CCC	AAA	GAC	AAT	GAC	ATG	CGCC	CTC	TAT	GAG	CTG	CAG	TCT	CCC	ACT	CAC	TCT	TTC	1103
151	GLY	THR	VAL	ARG	PRO	ILE	CYS	LEU	PRO	PHE	PHE	ASP	GLU	LEU	THR	PRO	ALA	THR	PRO	170	
1104	GGC	ACAG	T	CAG	GCC	CACT	GT	CTG	CCCT	TCT	TGAT	GAG	AGCT	CAC	TCC	AGC	CA	CCC	CA	1163	
171	LEU	TRP	ILE	ILE	GLY	TRP	GLY	PHE	THR	LYS	GLN	ASN	GLY	GLY	LYS	MET	SER	ASP	ILE	LEU	190
1164	CTT	GAT	CA	TTT	GAT	TGG	GCT	TTT	CA	GAG	CA	GAAT	TGG	AGG	AA	GAT	GTC	GAC	ATA	CTG	1223
191	LEU	GLN	ALA	SER	VAL	GLN	VAL	ILE	ASP	SER	THR	ARG	CYS	ASN	ALA	ASP	ALA	TYR	GLN	210	
1224	CTG	CAG	CGT	CAG	TCC	AGT	TCAT	TGAC	AGCAC	CAC	CGT	TGC	CAAT	GTC	AGAC	GAT	CGT	TAC	CG	1283	
211	GLY	GLU	VAL	THR	GLU	LYS	MET	MET	CYS	ALA	GLY	ILE	PRO	GLU	GLY	GLY	VAL	ASP	THR	CYS	230
1284	GGG	AGT	TC	ACC	GAG	AAG	ATG	ATG	TGT	GAG	GAT	CCCG	AA	GGG	GGG	TGT	GG	CACT	TGC	1343	
231	GLN	GLY	ASP	SER	GLY	GLY	PRO	LEU	MET	TYR	GLN	SER	ASP	GLN	TRP	HIS	VAL	VAL	GLY	ILE	250
1344	CAG	GTT	GAC	AGT	TGG	TGG	CCC	CTG	AT	GTT	ACC	AA	CTG	GACC	AGT	TGC	ATG	TGGT	GGG	CATC	1403
251	VAL	SER	TRP	GLY	TYR	GLY	CYS	GLY	GLY	PRO	SER	THR	PRO	GLY	VAL	TYR	THR	LYS	VAL	SER	270

Db	1404	GTTAGCTGGGGCTATGGCTCGGGGGCCCGAGCACCCCGAGGATATACACCAAGGTCTCA	1463
Qy	271	AlaryrieuAsnTrpIefyAsnValTrpLysAlaGluLeu	284
Db	1464	GCCTATCTCACTGGATCTACAATGTCTGGAAGGCTGAGCTG	1505
RESULT 10			
ABS74433			
ID	ABS74433 standard; cDNA; 2063 BP.		
XX	AC	ABS74433;	
XX	XX		
DT	10-DEC-2002 (first entry)		
XX			
DE	Human cDNA encoding secreted/transmembrane protein PRO1570.		
XX			
KW	Human; ss; gene; secreted protein; transmembrane protein; antirheumatic;		
KW	antiarthritic; osteopathic; sports-related joint problem;		
KW	articular cartilage defect; osteoarthritis; rheumatoid arthritis.		
XX			
OS	Homo sapiens.		
XX			
PN	US2002119130-A1.		
XX			
PD	29-AUG-2002.		
XX			
PF	06-DEC-2001; 2001US-0006867.		
XX			
PR	23-OCT-1997;	97US-063435P.	
PR	29-OCT-1997;	97US-064215P.	
PR	22-APR-1998;	98US-082797P.	
PR	29-APR-1998;	98US-083455P.	
PR	13-MAY-1998;	98US-085579P.	
PR	10-JUN-1998;	98US-088811P.	
PR	10-JUN-1998;	98US-088824P.	
PR	11-JUN-1998;	98US-088825P.	
PR	12-JUN-1998;	98US-088863P.	
PR	12-JUN-1998;	98US-089105P.	
PR	16-JUN-1998;	98US-089514P.	
PR	16-SEP-1998;	98WO-US19330.	
PR	08-MAR-1999;	99WO-US05028.	
PR	14-MAY-1999;	99WO-US10733.	
PR	02-JUN-1999;	99WO-US12252.	
PR	01-SEP-1999;	99WO-US20111.	
PR	15-SEP-1999;	99WO-US21030.	
PR	15-SEP-1999;	99WO-US21194.	
PR	22-DEC-1999;	99WO-US30720.	
PR	18-FEB-2000;	2000WO-US04341.	
PR	18-FEB-2000;	2000WO-US04342.	
PR	30-MAR-2000;	2000WO-US08439.	
PR	22-MAY-2000;	2000WO-US14042.	
PR	02-JUN-2000;	2000WO-US15264.	
PR	23-AUG-2000;	2000WO-US23522.	
PR	24-AUG-2000;	2000WO-US23328.	
PR	10-NOV-2000;	2000WO-US30873.	
PR	01-DEC-2000;	2000WO-US32378.	
PR	20-DEC-2000;	2000WO-US34956.	
PR	28-FEB-2001;	2001WO-US06520.	
PR	20-JUN-2001;	2001WO-US19692.	
PR	29-JUN-2001;	2001WO-US21066.	
PR	09-JUL-2001;	2001WO-US21735.	
XX			

(GETH) GENENTECH INC.

PI Eaton DL, Filvaroff E, Gerritsen ME, Goddard A, Godowski PJ;
 PI Grimaldi JC, Gurney AL, Watanabe CK, Wood WI;

DR WPI; 2002-731348/79.

P-FSDE; ABG35906.
XX
PT New isolated secreted and transmembrane PRO polypeptide useful for
PT modulating biological activity of a cell, or for treating
PT sports-related joint problems, osteoarthritis or rheumatoid arthritis

Claim 2; Fig 111; 399pp; English.

The invention relates to an isolated secreted and transmembrane PRO polypeptide having 80 % sequence identity to a sequence appearing as ABG9581-ABG9594 or their associated signal peptide, or a sequence of an extracellular domain of the proteins with their associated signal peptide or lacking its associated signal peptide. Also included are the nucleic acids encoding the proteins, vectors, host cells, fusion proteins and antibodies which specifically bind to the proteins. The proteins are useful for detecting a polypeptide designated as A, B, C or D in a sample suspected of containing an A, B, C or D polypeptide, by contacting the sample with a polypeptide designated as E, F, G, H or I (or vice versa) and determining the formation of a A/E, B/F, B/G, C/H or D/I polypeptide conjugate in the sample, where the formation of the conjugate is indicative of the presence of an A, B, C or D polypeptide in the sample, where A is a PRO10272 polypeptide, B is a PRO20110 polypeptide, C is a PRO10096 polypeptide, D is a PRO19760 polypeptide, E is a PRO5801 polypeptide, F is a PRO1 polypeptide, G is a PRO20040 polypeptide, H is a PRO20233 polypeptide and I is a PRO1890 polypeptide. The sample comprises a cell suspected of expressing the A, B, C or D polypeptide. The E, F, G, H or I polypeptide is labeled with a detectable label or is attached to a solid support. The proteins are useful for linking a bioactive molecule to a cell expressing a polypeptide designated as A, B, C or D or E, F, G, H or I. The bioactive molecule is a toxin, a radiolabel or an antibody. The bioactive molecule causes death of the cell. A, B, C, D, E, F, G, H, or I, or antibodies against them are useful for modulating a biological activity of a cell expressing a polypeptide designated as A, B, C or D or E, F, G, H, or I. The cell is killed. The proteins are useful for identifying agonists or antagonists, for the preparation of a medicament useful in the treatment of a condition which is responsive to the proteins, as molecular weight markers for protein electrophoresis purposes, and as therapeutic agents for treating sports-related joint problems, articular cartilage defects, osteoarthritis or rheumatoid arthritis. Nucleic acids encoding the proteins are useful as hybridisation probes, in chromosome and gene mapping, in the generation of anti-sense RNA and DNA, for the preparation of the proteins, to generate transgenic or knockout animals which are useful in the development and screening of therapeutic useful reagents, for chromosome identification, and in gene therapy. The antibody is useful as a therapeutic agent, and in gene assay and for affinity purification of the protein from recombinant cell culture natural sources. The present sequence encodes a novel secreted or transmembrane protein of the invention.

Sequence 2063 BP; 477 A; 591 C; 576 G; 419 T; 0 other;

alignment Scores:					
red. No.:	2.65e-120	Length:	2063		
core:	1266.00	Matches:	238		
percent Similarity:	98.129%	Conservative:	2		
best Local Similarity:	97.44%	Mismatches:	4		
nery Match:	79.12%	Indels:	0		
B:	24	Gaps:	0		
 IS-09-607-745-9 (1-292) x ABS74433 (1-2063)					
y	51	LysIleValGlyGlyTyrAlaLeuAspValAspSerTrpProTrpGlnValSerIlecln	70		
b	804	CGTGTTGGGTGGGAGGAGGCCTCTGGGATCTTGGCCCTTGCAGTCACGATCCAG	863		
y	71	TyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThrAla	90		
b	864	TACGACAAACAGACGCTCTGTGGAGGAGCATCTTGCACCCCACATGGGTCTCTACGGCA	923		
y	91	AlaHisCysPheArgIlysthisThrAspValPheAsnTrpLysValArgAlaGlySerAsp	110		
b	924	GCCCACTGCTTCAGGAACAATACCGATGTGTTCAACTGGAAGTGGGGCAGGCTCAGAC	983		
y	111	LysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleIleIleclnPhaAsnPro	130		
b	984	AAACTGGGCACTGCCATCCCTGGCTGTGGCCAACATCATCATTAATTCAACCCC	1043		

QY	131	MetTyrProIysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPheSer	150
DB	1044	ATGTACCCCAAGACAAATGACATCGCCCTCATGAAGCTGCAGTTCCCACTCACTTTCTCA	1103
QY	151	GlyThrValArgProIleCysLeuProPhePheAspGluLeuThrProAlaThrPro	170
DB	1104	GGCACAGTCAGGCCCAATCTGTCTGCCCTCTTTTGATGAGAGCTCACTCAGCACCCCA	1163
QY	171	LeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyGlyLysMetSerAspIleLeu	190
DB	1164	CYCTGGATCATTTGGATGGGCTTTTACGAGCAGAATGGAGGGAAGATGCTGCACACTG	1223
QY	191	LeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAlaTyrGln	210
DB	1224	CTCAGCGCTCAGTCCCACTATTGACAGCACACCGTGCAATGCAGACGATGGTACCAG	1283
QY	211	GlyGlnValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThrCys	230
DB	1284	GGGGAAGTACCAGGAAGATGATGTGCAGGATCCCGAAGGGGTGGACACCTGC	1343
QY	231	GlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGlyIle	250
DB	1344	CAGGGTGACAGTGTGGGCCCTGATGTACCAATCTGACCAGTGGCATGTGGTGGCATC	1403
QY	251	ValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyValTyrThrLysValSer	270
DB	1404	GTTAGCTGGGGCTATGGCTCGCGGGGCCGAGCACCCAGGAGTATACACCAAGGTCTCA	1463
QY	271	AlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu	284
DB	1464	GCCTATCTCACTGGATCTACAAATGCTGGAGGGCTGAGCTG	1505
RESULT 11			
ABK11090			
ID	ABK11090 standard; cDNA; 2063 BP.		
XX			
AC	ABK11090;		
XX			
XX	05-JUN-2002 (first entry)		
XX	CDNA encoding tumour-associated antigenic target protein, TAT135.		
DE			
XX	TAT135; Tumour-associated Antigenic Target; tumour;		
KW	breast cancer; colorectal cancer; lung cancer; ovarian cancer;		
KW	central nervous system cancer; liver cancer; bladder cancer; melanoma;		
KW	pancreatic cancer; leukaemia; gene therapy; gene; ss.		
XX			
OS	Homo sapiens.		
XX			
Key	Location/Qualifiers		
FH	210..1508		
FT	/*tag= a		
FT	/product= "TAT135"		
FT	/note= "tumour-associated antigenic target"		
XX			
PN	WO200216429-A2.		
XX			
PD	28-FEB-2002.		
XX			
XX	22-JUN-2001; 2001WO-US20118.		
PF			
XX			
PR	24-AUG-2000; 2000WO-US23328.		
PR	26-SEP-2000; 2000US-235451P.		
PR	01-DEC-2000; 2000WO-US32678.		
PR	28-FEB-2001; 2001WO-US06520.		
PR	01-MAR-2001; 2001WO-US06666.		
XX			
PA	(GETH) GENENTECH INC.		
XX			
XX	Goddard A, Godowski PJ, Gurney AL, Hillan KJ, Polakis P, Smith V;		
PI	Wood WI, Wu TD, Zhang Z;		
XX			
DR	WPI: 2002-280917/32.		

R 20-DEC-2000; 2000WO-US34956.
R 28-FEB-2001; 2001WO-US06520.
R 01-JUN-2001; 2001WO-US17800.
R 20-JUN-2001; 2001WO-US19692.
R 29-JUN-2001; 2001WO-US21066.
R 09-JUL-2001; 2001WO-US21735.
R 29-AUG-2001; 2001WO-US27099.
R 18-SEP-1997; 97US-059263P.
R 18-SEP-1997; 97US-059266P.
R 17-OCT-1997; 97US-062250P.
R 21-OCT-1997; 97US-063486P.
R 24-OCT-1997; 97US-063120P.
R 24-OCT-1997; 97US-063121P.
R 28-OCT-1997; 97US-063540P.
R 28-OCT-1997; 97US-063541P.
R 28-OCT-1997; 97US-063544P.
R 28-OCT-1997; 97US-063564P.
R 29-OCT-1997; 97US-063734P.
R 31-OCT-1997; 97US-063870P.
R 31-OCT-1997; 97US-064103P.
R 13-NOV-1997; 97US-065311P.
R 21-NOV-1997; 97US-066120P.
R 24-NOV-1997; 97US-066466P.
R 24-NOV-1997; 97US-066772P.
R 11-DEC-1997; 97US-069335P.
R 12-DEC-1997; 97US-069425P.
R 17-DEC-1997; 97US-069870P.
R 18-DEC-1997; 97US-068017P.
R 10-MAR-1998; 98US-077450P.
R 11-MAR-1998; 98US-077632P.
R 11-MAR-1998; 98US-077649P.
R 20-MAR-1998; 98US-078886P.
R 20-MAR-1998; 98US-078939P.
R 27-MAR-1998; 98US-079864P.
R 27-MAR-1998; 98US-079786P.
R 31-MAR-1998; 98US-080107P.
R 01-APR-1998; 98US-080327P.
R 01-APR-1998; 98US-080333P.
R 08-APR-1998; 98US-081049P.
R 08-APR-1998; 98US-081070P.
R 09-APR-1998; 98US-081195P.
R 15-APR-1998; 98US-081838P.
R 21-APR-1998; 98US-082568P.
R 21-APR-1998; 98US-082669P.
R 22-APR-1998; 98US-082704P.
R 22-APR-1998; 98US-082797P.
R 28-APR-1998; 98US-083322P.
R 29-APR-1998; 98US-083495P.
R 29-APR-1998; 98US-083496P.
R 29-APR-1998; 98US-083499P.
R 29-APR-1998; 98US-083559P.
R 05-MAY-1998; 98US-084366P.
R 06-MAY-1998; 98US-084414P.
R 07-MAY-1998; 98US-084639P.
R 07-MAY-1998; 98US-084640P.
R 07-MAY-1998; 98US-084643P.
R 15-MAY-1998; 98US-085579P.
R 15-MAY-1998; 98US-085580P.
R 15-MAY-1998; 98US-085582P.
R 18-MAY-1998; 98US-085700P.
R 18-MAY-1998; 98US-086023P.
R 22-MAY-1998; 98US-086392P.
R 22-MAY-1998; 98US-086486P.
R 28-MAY-1998; 98US-087098P.
R 28-MAY-1998; 98US-087208P.
R 02-JUN-1998; 98US-087609P.
R 02-JUN-1998; 98US-087759P.
R 03-JUN-1998; 98US-087827P.
R 04-JUN-1998; 98US-088025P.
R 04-JUN-1998; 98US-088028P.
R 04-JUN-1998; 98US-088029P.
R 04-JUN-1998; 98US-088033P.
PR 04-JUN-1998; 98US-088326P.
PR 05-JUN-1998; 98US-088167P.
PR 05-JUN-1998; 98US-088202P.
PR 05-JUN-1998; 98US-088212P.
PR 05-JUN-1998; 98US-088217P.
PR 09-JUN-1998; 98US-088655P.
PR 10-JUN-1998; 98US-088722P.
PR 10-JUN-1998; 98US-088738P.
PR 10-JUN-1998; 98US-088740P.
PR 10-JUN-1998; 98US-088811P.
PR 10-JUN-1998; 98US-088824P.
PR 10-JUN-1998; 98US-088825P.
PR 10-JUN-1998; 98US-088826P.
PR 11-JUN-1998; 98US-088861P.
PR 11-JUN-1998; 98US-088863P.
PR 11-JUN-1998; 98US-088876P.
PR 12-JUN-1998; 98US-089090P.
PR 12-JUN-1998; 98US-089105P.
PR 16-JUN-1998; 98US-089512P.
PR 16-JUN-1998; 98US-089514P.
PR 17-JUN-1998; 98US-089538P.
PR 17-JUN-1998; 98US-089598P.
PR 17-JUN-1998; 98US-089653P.
PR 18-JUN-1998; 98US-089908P.
PR 19-JUN-1998; 98US-089952P.
PR 22-JUN-1998; 98US-090246P.
PR 22-JUN-1998; 98US-090252P.
PR 24-JUN-1998; 98US-090429P.
PR 24-JUN-1998; 98US-090435P.
PR 24-JUN-1998; 98US-090444P.
PR 24-JUN-1998; 98US-090461P.
PR 24-JUN-1998; 98US-090535P.
PR 25-JUN-1998; 98US-090540P.
PR 25-JUN-1998; 98US-090676P.
PR 25-JUN-1998; 98US-090688P.
PR 25-JUN-1998; 98US-090690P.
PR 25-JUN-1998; 98US-090694P.
PR 25-JUN-1998; 98US-090695P.
PR 25-JUN-1998; 98US-090696P.
PR 26-JUN-1998; 98US-090862P.
PR 26-JUN-1998; 98US-090863P.
PR 26-JUN-1998; 98US-091010P.
PR 01-JUL-1998; 98US-091359P.
PR 01-JUL-1998; 98US-091544P.
PR 02-JUL-1998; 98US-091478P.
PR 02-JUL-1998; 98US-091486P.
PR 02-JUL-1998; 98US-091626P.
PR 02-JUL-1998; 98US-091628P.
PR 02-JUL-1998; 98US-091632P.
PR 24-JUL-1998; 98US-094006P.
PR 04-AUG-1998; 98US-095282P.
PR 10-AUG-1998; 98US-095998P.
PR 10-AUG-1998; 98US-096012P.
PR 17-AUG-1998; 98US-096757P.
PR 17-AUG-1998; 98US-096766P.
PR 17-AUG-1998; 98US-096867P.
PR 17-AUG-1998; 98US-096891P.
PR 17-AUG-1998; 98US-096897P.
PR 18-AUG-1998; 98US-096949P.
PR 18-AUG-1998; 98US-096959P.
PR 18-AUG-1998; 98US-097022P.
PR 26-AUG-1998; 98US-097952P.
PR 26-AUG-1998; 98US-097954P.
PR 26-AUG-1998; 98US-097955P.
PR 26-AUG-1998; 98US-097971P.
PR 26-AUG-1998; 98US-097974P.
PR 26-AUG-1998; 98US-098014P.
PR 01-SEP-1998; 98US-098716P.
PR 01-SEP-1998; 98US-098723P.
PR 02-SEP-1998; 98US-098803P.
PR 02-SEP-1998; 98US-098821P.

```

R 02-SEP-1998; 98US-098843P.
R 09-SEP-1998; 98US-099602P.
R 10-SEP-1998; 98US-099741P.
R 10-SEP-1998; 98US-099754P.
R 10-SEP-1998; 98US-099763P.
R 10-SEP-1998; 98US-099812P.

Alignment Scores:
Score: 2.65e-120 2063
Length: 1266.00
Matches: 228
Conservative: 2
Mismatch: 4
Indels: 0
Gaps: 0

us-09-607-745-9 (1-292) x ACA57847 (1-2063)
51 LysIleValGlyGlyTyrAlaLeuAspValAspSerTrpProTrpGlnValSerIleGln 70
804 CQTGTGGTGGTGGGAGGAGGCTCTGTGGATCTTGGCCCTTGGCAGGTGACATCCAG 863
71 TyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThrAla 90
864 TAGCAAAACAGACAGCTCTGTGGAGGAGCATCTCTGACCCCACTGGTCTCAGGGCA 923
91 AlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySerAsp 110
924 GCCCACTGTTCTCAGGAACATACCGATGTTCACTGGAGGTGGCGCAGGCTCAGAC 983
111 LysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleGluPheAsnPro 130
984 AAATGGGAGCTTCCATCCCTGGCTGGCCCAAGATCATCATCTTGAATCAACCC 1043
131 MetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPheSer 150
1044 ATGTACCCCAACACATGATGATGCGCCCTCATGAGTGCAGTCCACTCACTTTCTCA 1103
151 GlyThrValArgProIleCysLeuProPhePheAspGluGluLeuThrProAlaThrPro 170
1104 GGACAGTCAGGCCCCATCTGTGCCCCCTTCTTGTAGAGAGCTCACTCAGCCACCCCA 1163
171 LeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyGlyLysMetSerAspIleLeu 190
1164 CTCTGATCATTTGGATGGGCTTTACAGAGCAGATGGAGGAGATGCTGACATACG 1223
191 LeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAlaTrpGln 210
1224 CTGCAGGCTCAGTCCAGGTCATTCAGACACACGCTGCAATGCAGACGATGCTACCA 1283
211 GlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThrCys 230
1284 GGGGAAGTCACCGAGAAGATGATGTGTGAGGATCCCGAAGGGGGTGTGACACCTGC 1343
231 GlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGlyIle 250
1344 CAGGTGACAGTGTGGGCCCTTGATGTACCAATCTGACCACTGGCATGTGTGGGCATC 1403
251 ValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyValTrpThrLysValSer 270
1404 GTTAGTGGGGTATGGTGGGGGGCCCGAGCACCCAGGAGTATACACCAAGGTCTCA 1463
271 AlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu 284
1464 GCTATCTCACTGGATCTACATGATCTGTGAAGGCTGAGCTG 1505

```

RESULT 13

ACA58865

D ACA58865 standard; cDNA; 2063 BP.

X ACA58865;

X ACA58865;

X ACA58865;

10-JUN-2003 (first entry)

DE cDNA encoding human secreted polypeptide PROi570.

XX Human; ss; gene; gene therapy; tumour; cancer.

XX Homo sapiens.

XX US2003013855-A1.

XX 16-JAN-2003.

XX 03-MAY-2002; 2002US-0063616.

XX 30-DEC-1998; 98KR-0062142.

XX 08-MAR-1999; 99WO-US05028.

XX 14-MAY-1999; 99WO-US10733.

XX 30-DEC-1999; 99WO-US31274.

XX 18-FEB-2000; 2000WO-US04341.

XX 01-MAR-2000; 2000WO-US05601.

XX 02-MAR-2000; 2000WO-US05841.

XX 21-MAR-2000; 2000WO-US07532.

XX 22-MAY-2000; 2000WO-US14042.

XX 02-JUN-2000; 2000WO-US15264.

XX 24-AUG-2000; 2000WO-US23328.

XX 10-NOV-2000; 2000WO-US30873.

XX 01-DEC-2000; 2000WO-US32678.

XX 20-DEC-2000; 2000WO-US34956.

XX 28-FEB-2001; 2001WO-US06520.

XX 01-JUN-2001; 2001WO-US17800.

XX 14-MAY-1999; 99US-0311832.

XX 25-AUG-1999; 99US-0380137.

XX 25-AUG-1999; 99US-0380138.

XX 25-AUG-1999; 99US-0380139.

XX 25-AUG-1999; 99US-0380142.

XX 15-SEP-1999; 99US-0397342.

XX 18-OCT-1999; 99US-0403297.

XX 12-NOV-1999; 99US-0423844.

XX 22-AUG-2000; 2000US-0644848.

XX 18-SEP-2000; 2000US-0664610.

XX 18-SEP-2000; 2000US-0665350.

XX 08-NOV-2000; 2000US-0709238.

XX 20-DEC-2000; 2000US-0747259.

XX 22-MAR-2001; 2001US-0816744.

XX 10-MAY-2001; 2001US-0854208.

XX 10-MAY-2001; 2001US-0854280.

XX 30-MAY-2001; 2001US-0870574.

XX 05-JUN-2001; 2001US-0874503.

XX 29-JUN-2001; 2001US-0869599.

XX 18-JUL-2001; 2001US-0908827.

XX 06-DEC-2001; 2001US-0006867.

XX (GETH) GENENTECH INC.

XX Eaton DL, Filvaroff E, Gerritsen ME, Goddard A, Godowski PJ;

XX Grimaldi JC, Gurney AL, Watanabe CK, Wood WL;

XX WPI; 2003-330485/31.

XX P-PSDB; ABU71561.

XX New isolated antibody specifically binding a PRO polypeptide, useful

XX for the preparation of a medicament for treating disorders with the

XX aberrant expression or activity of the PRO polypeptide, such as tumor

XX conditions and cancer -

XX Example 4; Page 184-185; 406pp; English.

XX The invention relates to an antibody that binds to a polypeptide with a

XX fully defined sequence given in the specification. The methods and

XX compositions (containing antibodies that specifically bind a PRO

XX polypeptide) of the present invention are useful for the preparation of a

XX medicament for the treatment of disorders associated with the aberrant

XX expression or activity of the PRO polypeptide, such as tumour conditions

XX and cancer. They can also be used to generate transgenic or knockout

XX animals useful in the development and screening of therapeutically useful

QY 71 TyrAspLysGlnHisValCysGlyGlySerIleIeuAspProHisTrpValLeuThrAla 90
Db 864 TACGACAAACAGACGCTGTGGAGGGAGCATCTTGGACCCCACTGGGTCTTCACGCCA 923

Qy 71 TyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValIleuThrAla 90
Db 864 TACGACAAACAGTACGTCCTGTGACGCGACATCGTGGACCCCCATGCGCTCTCTACCGCA 932

91 AlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySerAsp 110
|||
924 GGCACCTGCTTCAGGAAGACATACCGATGTGTTCAACTGGAGGTGGGCGCTCAGAC 983
|||
111 LysLeuGlySerPheProSerLeuAlaValAlaValLeuLeuLeuLeuLeuLeuLeu 130
|||
984 AAACCTGGGAGCTTCCATCCCTGGGTGGGCAAGATCATCATTAATGAATCAACCC 1043
|||
131 MetTyrProLysAspAsnAspAlaLeuMetLysLeuGluPheProLeuThrPheSer 150
|||
1044 ATGTACCCCAAGACATGACATGGCCCTCATGAAGCTGAGTTCCTCACTTTCTCA 1103
|||
151 GlyThrValArgProLysCysLeuProPhePheAspGluGluLeuThrProAlaThrPro 170
|||
1104 GGCACAGTCAGGCCCATCTGTCTGCCCTTCTTTGATGAGGAGCTCACTCCAGCCACCCCA 1163
|||
171 LeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyGlyLysMetSerAspIleLeu 190
|||
1164 CTCTGGATCATTTGGATGGGCTTTACGAAGCAGATGGAGGAGATGCTGACATACGT 1223
|||
191 LeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAlaThrGln 210
|||
1224 CTGAGGGCTCAGTCCAGGTGATTTGACAGCAGCGTGCAATGACAGATGCGGTACCG 1283
|||
211 GlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThrCys 230
|||
1284 GGGGAAGTCACCGAGAAGATGATGTGTGAGGAGCATCCCGAAGGGGGTGTGACACCTGC 1343
|||
231 GlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGlyIle 250
|||
1344 CAGGGTGACAGTGTGGGCCCTGTATGTACCAATCTGACCATGGCATGTGTGGGCATC 1403
|||
251 ValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyValThrLysValSer 270
|||
1404 GTTAGCTGGGCTATGGTGTGGGCGGCGGAGCACCAGGAGTATACACCAAGTCTCA 1463
|||
271 AlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu 284
|||
1464 GCCTATCTCACTGATCTACAATGTCTGGAAGGCTGAGCTG 1505
|||

RESULT 15

ACA63428
ACA63428 standard; cDNA; 2063 BP.
ACA63428;
13-JUN-2003 (first entry)
cDNA encoding human PRO polypeptide #56.
Human; PRO polypeptide; secreted and transmembrane protein;
anti-PRO antibody; diagnostic assay; gene expression; gene, ss.

Homo sapiens.

US2003023042-A1.

30-JAN-2003.

01-MAY-2002; 2002US-0063502.

06-DEC-2001; 2001US-0006867.

(GETH) GENENTECH INC.

Eaton DL, Filvaroff E, Gerritsen ME, Goddard A, Godowski PJ;
Grimaldi JC, Gurney AL, Watanabe CK, Wood WI;

WPI; 2003-331484/31.

P-P8DB; ABU72164.

Novel monoclonal antibody that binds to secreted and transmembrane

PT polypeptide, useful for detecting and purifying the polypeptide and
PR also for treating conditions responsive to the antibody

PS Disclosure; Fig 111; 408pp; English.

XX The present invention relates to the isolation of novel human PRO
CC polypeptides, and the polynucleotide sequences encoding them. The
CC PRO polypeptides are secreted and transmembrane proteins. The PRO
CC polypeptides and polynucleotides are useful for preparing a
CC medicament useful in the treatment of a condition responsive to
CC anti-PRO antibody. Anti-PRO antibodies are useful in diagnostic
CC assays for PRO, by detecting its expression in specific cells,
CC tissues or serum, and for affinity purification of PRO from
CC recombinant cell culture or natural purifications. ACA63373-ACA63456
CC represent cDNA sequences encoding the human PRO polypeptides of
CC the invention.

SQ Sequence 2063 BP; 477 A; 591 C; 576 G; 419 T; 0 other;

Alignment Scores:

Pred. No.:	2,65e-120	Length:	2063
Score:	1266.00	Matches:	228
Percent Similarity:	98.29%	Conservative:	2
Best Local Similarity:	97.44%	Mismatches:	4
Query Match:	79.12%	Indels:	0
DB:	25	Gaps:	0

US-09-607-745-9 (1-292) x ACA63428 (1-2063)

QY	51	LysIleValGlyGlyTyrAlaLeuAspValAspSerTrpProTrpGlnValSerIleGln	70
DB	804	CGTGTGGTGGGTGGGAGGAGGCGCTCTGTGGATTCTTGGCCCTTGGCAGTCAGCATCCAG	863
QY	71	TyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThrAla	90
DB	864	TACGACAAACGACGCTCTGTGGAGGAGCATCTTGGACCCCTCTGGTCTCTCAGGCA	923
QY	91	AlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySerAsp	110
DB	924	GCCCACTGCTTCAGGAACATACCGATGTGTTCACCTGGAAGTGGCGGAGGCTCAGAC	983
QY	111	LysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleIleGluPheAsnPro	130
DB	984	AAACTGGGCGCTTCCCATCCCTGGCTGTGGCCAGATCATCATCAATTAATCAACCC	1043
QY	131	MetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPheSer	150
DB	1044	ATGTACCCCAAGACATGACATGGCCCTCATGAAGTGGCAGTTCCTCACTTTCTCA	1103
QY	151	GlyThrValArgProLysCysLeuProPhePheAspGluGluLeuThrProAlaThrPro	170
DB	1104	GGCAGTCAGGCCCATCTGTCTGCCCTTCTTTGATGAGGAGCTCACTCCAGCCACCCCA	1163
QY	171	LeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyGlyLysMetSerAspIleLeu	190
DB	1164	CTCTGGATCATTTGGATGGGCTTTACGAAGCAGATGGAGGAGATGCTGACATACGT	1223
QY	191	LeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAlaThrGln	210
DB	1224	CTGAGGGCTCAGTCCAGGTGATTTGACAGCAGCGTGCAATGACAGATGCGGTACCG	1283
QY	211	GlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThrCys	230
DB	1284	GGGGAAGTCACCGAGAAGATGATGTGTGAGGAGCATCCCGAAGGGGGTGTGACACCTGC	1343
QY	231	GlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGlyIle	250
DB	1344	CAGGGTGACAGTGTGGGCCCTGTATGTACCAATCTGACCATGGCATGTGTGGGCATC	1403
QY	251	ValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyValThrLysValSer	270
DB	1404	GTTAGCTGGGCTATGGTGTGGGCGGCGGAGCACCAGGAGTATACACCAAGTCTCA	1463

Y 271 AlaTyrLeuAsnTrpIleTyrAsnValTyrLysAlaGluLeu 284
b ||||||||||||||||||||||||||||||||||||||||
1464 GCCTATCTCAACTGGATCTACAAATGTCGAGAGGCTGAGCTG 1505

Search completed: November 23, 2003, 08:07:00
Job time : 280.081 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

M protein - nucleic search, using frame_plus_p2n model

run on: November 23, 2003, 07:53:55 ; Search time 58.641 Seconds
(without alignments)
2197.846 Million cell updates/sec

title: US-09-607-745-9
perfect score: 1600
sequence: 1 MDSKSSQKSRLLLLLVSN.....LNWVNVKAEISRHHHHH 292

scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

searched: 569978 seqs, 220691566 residues

total number of hits satisfying chosen parameters: 1139956

minimum DB seq length: 0
maximum DB seq length: 2000000000

post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

command line parameters:

MODEL=frame+p2n.model -DEV=xlh
MODEL2=/usr2/ptdptool/US09607745/runat_21112003_144346_22244/app_query.fasta_1.1038
-2=/cgn2_1/USPTO.spool/US09607745 -SUFTX=rni -MINMATCH=0.1 -LOOPCL=0
-DB=Issued Patents NA -QFMT=fastap -START=1 -ENDS=1 -MATRIX=blomum62 -TRANS=human40.cdi
-COOPEXT=0 -UNITS=bits -START=1 -ENDS=1 -MATRIX=blomum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
US09607745 @CGN_1_1_76 @runat_21112003_144346_22244 -NCPU=6 -ICPU=3
NO MWAP -LARGQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-XGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

database :

Issued Patents NA.*
1: /cgn2_6/ptodata/2/ina/5A COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PCTUS COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1266	79.1	2038	3	US-09-008-271A-18
2	1266	79.1	2079	4	US-09-656-002-1
3	765	47.8	1130	4	US-09-387-375-8
4	754.5	47.2	1130	4	US-09-386-653A-8
5	746	46.6	1169	4	US-09-386-642-7
6	740.5	46.3	1166	4	US-09-386-629-2
7	672.5	42.0	1052	4	US-09-386-642-10
8	645	40.3	1049	4	US-09-386-642-9
9	601.5	37.6	1103	4	US-09-386-642-59
10	593	37.1	1142	4	US-09-386-642-8
11	574	35.9	1037	4	US-09-386-642-60
12	526.5	32.9	1460	4	US-09-370-838-80

13	526.5	32.9	1517	1	US-08-508-448C-15	Sequence 15, Appl
14	526.5	32.9	2790	4	US-09-370-838-79	Sequence 79, Appl
15	524.5	32.8	1462	4	US-09-370-838-55	Sequence 55, Appl
16	522.5	32.7	696	1	US-08-508-448C-24	Sequence 24, Appl
17	510.5	31.9	901	1	US-08-508-448C-9	Sequence 9, Appl
18	508	31.8	2479	3	US-09-342-749-29	Sequence 29, Appl
19	508	31.8	2479	4	US-09-691-840-29	Sequence 29, Appl
20	506	31.6	1479	3	US-09-342-749-1	Sequence 1, Appl
21	506	31.6	1479	4	US-09-691-840-1	Sequence 1, Appl
22	505.5	31.6	1615	4	US-09-820-002-1	Sequence 1, Appl
23	505.5	31.6	1783	3	US-09-510-738A-188	Sequence 188, App
24	505.5	31.6	1783	4	US-09-861-966-188	Sequence 188, App
25	505.5	31.6	2363	4	US-09-742-703-3	Sequence 3, Appl
26	502	31.4	2413	3	US-09-518-046-1	Sequence 1, Appl
27	497	31.1	1077	3	US-08-807-151-2	Sequence 2, Appl
28	497	31.1	1077	4	US-09-478-957-2	Sequence 2, Appl
29	491.5	30.7	1605	2	US-09-000-846-1	Sequence 1, Appl
30	482	30.1	959	4	US-09-023-942A-25	Sequence 25, Appl
31	478.5	29.9	2544	3	US-09-518-046-3	Sequence 3, Appl
32	473	29.6	2416	3	US-09-261-416-1	Sequence 1, Appl
33	456.5	28.5	1100	4	US-09-023-942A-5	Sequence 5, Appl
34	455	28.4	1613	4	US-09-387-375-1	Sequence 1, Appl
35	453.5	28.3	897	2	US-08-956-267A-1	Sequence 1, Appl
36	451.5	28.2	1081	3	US-09-008-271A-15	Sequence 15, Appl
37	450.5	28.2	1094	4	US-09-023-942A-3	Sequence 3, Appl
38	444	27.8	3147	2	US-09-027-337-1	Sequence 1, Appl
39	444	27.8	3147	4	US-09-644-600-1	Sequence 1, Appl
40	444	27.8	3147	4	US-09-644-600-18	Sequence 18, Appl
41	443.5	27.7	825	3	US-09-120-582-1	Sequence 1, Appl
42	442.5	27.7	980	4	US-09-023-942A-30	Sequence 30, Appl
43	442.5	27.7	1110	4	US-09-386-653A-1	Sequence 1, Appl
44	437.5	27.3	1212	4	US-09-620-312D-431	Sequence 431, App
45	431.5	27.0	1225	4	US-09-734-675-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1

US-09-008-271A-18

; Sequence 18, Application US/09008271A

; Patent No. 6203979

; GENERAL INFORMATION:

; APPLICANT: Bandman, Olga

; Hillman, Jennifer L.

; Yue, Henry

; Guegler, Karl J.

; Corley, Neil C.

; Tang, Tom Y.

; Shah, Purvi

; TITLE OF INVENTION: HUMAN PROTEASE MOLECULES

; NUMBER OF SEQUENCES: 24

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Incyte Pharmaceuticals, Inc.

; STREET: 3174 Porter Dr.

; CITY: Palo Alto

; STATE: CA

; COUNTRY: USA

; ZIP: 94304

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; SOFTWARE: FastSeq for Windows Version 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/008,271A

; FILING DATE: 16-Jan-1998

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: <Unknown>

; FILING DATE: <Unknown>

; ATTORNEY/AGENT INFORMATION:

; NAME: Mohan-Peterson, Sheela

; REGISTRATION NUMBER: 41,201

; REFERENCE/DOCKET NUMBER: PF-0458 US

TELECOMMUNICATION INFORMATION:

TELEPHONE: 650-855-0555

TELEFAX: 650-845-4166

INFORMATION FOR SEQ ID NO: 18:

SEQUENCE CHARACTERISTICS:

LENGTH: 2038 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

IMMEDIATE SOURCE:

LIBRARY: COLN00113

CLONE: 1337018

SEQUENCE DESCRIPTION: SEQ ID NO: 18:

us-09-608-271A-18

Alignment Scores:
Pred. No.: 5,58e-130 Length: 2038
Score: 1266.00 Matches: 228
Percent Similarity: 98.29% Conservative: 2
Best Local Similarity: 97.44% Mismatches: 4
Query Match: 79.12% Indels: 0
DB: 3 Gaps: 0

US-09-607-745-9 (1-292) x US-09-008-271A-18 (1-2038)

51 LysileValGlyGlyTyrAlaLeuAspValAspSerTrpProTrpGlnValSerIleGln 70
803 CQTGTGGTGGGTGGGAGGAGGCTCTGTGGATTCCTGGCTTGGCAGTCCAGTCCAG 862
71 TyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThra 90
863 TAGCAAAACAGACAGCTGTGTGGAGGAGATCTCTGGACCCACCTGGGTCTCCAGGCA 922
91 AlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySerAsp 110
923 GCCACTGCTTCAGAAACATACCGATGTTCACTGGAGGTGGGCGAGCTCAGAC 982
111 LysLeuGlySerPheProSerIleuAlaValAlaLysIlellellellelleGluPheAsnPro 130
983 AAACCTGGGAGCTTCCATCCCTGGTGTGGCCAGATCATCATCATCATCATCATCAACCC 1042
131 MetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThraPheSer 150
1043 ATGTACCCCAAGACAAATGACATCGCCCTCATGAAGCTGAGTCCCATCTCACTTCTCA 1102
151 GlyThrValArgProIleCysLeuProPhePheAspGluGluLeuThraProAlaThrPro 170
1103 GGCACAGTCAGGCCCATCTGTCTGCCCTTCTTGTAGAGGAGCTCACTCCAGCCACCCCA 1162
171 LeuTrpIlellelleGlyTrpGlyPheThrLysGlnAsnGlyGlyLysMetSerAspIleLeu 190
1163 CTTCTGGATCATTTGGATGGGGCTTTACGAGCAGATGGAGGGAAGATGCTGCACATCTG 1222
191 LeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAspAlaTyrGln 210
1223 CTGCAGGCTCAGTCCAGTCAATGACACACAGGTGCAATGACACAGTCCGATCCAG 1282
211 GlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThrCys 230
1283 GGGGAAGTCACCGAGAAGATGATGTGTGCAGGATCCCGGAGGGGGGTGTGCACCTCTGC 1342
231 GlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGlyIle 250
1343 CAGGGTGACAGTGTGGTGGCCCTTGATGTACCAATCTGACACAGTGGCATGTGTGGGCATC 1402
251 ValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyValThrLysValSer 270
1403 GTTAGCTGGGCTATGGCTCGGGGGGCCCGAGCACCCAGGAGATATACACCAAGGTCTCA 1462
271 AlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu 284
1463 GCCTATCTCACTGGATCTACATGATGCTGGAGGCTGAGCTG 1504

RESULT 2

US-09-656-002-1
; Sequence 1, Application US/09656002
; Patent No. 6455668
; GENERAL INFORMATION:
; APPLICANT: Mack, David
; APPLICANT: Wilson, Keith
; APPLICANT: Gish, Kurt
; TITLE OF INVENTION: NOVEL METHODS OF DIAGNOSING COLORECTAL CANCER, COMPOSITIONS, AND
; FILE REFERENCE: A-69108/DJB/JJD/AMS
; CURRENT APPLICATION NUMBER: US/09/656,002
; CURRENT FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: US 09/525,993
; PRIOR FILING DATE: 2000-03-15
; PRIOR APPLICATION NUMBER: US 09/493,444
; PRIOR FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: PCT/US 00/07044
; PRIOR FILING DATE: 2000-03-15
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 2079
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-656-002-1

Alignment Scores:
Pred. No.: 5,74e-130 Length: 2079
Score: 1266.00 Matches: 228
Percent Similarity: 98.29% Conservative: 2
Best Local Similarity: 97.44% Mismatches: 4
Query Match: 79.12% Indels: 0
DB: 4 Gaps: 0

US-09-607-745-9 (1-292) x US-09-656-002-1 (1-2079)

51 LysileValGlyGlyTyrAlaLeuAspValAspSerTrpProTrpGlnValSerIleGln 70
818 CQTGTGGTGGGTGGGAGGAGGCTCTGTGGATTCCTGGCTTGGCAGTCCAGTCCAG 877
71 TyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThra 90
878 TAGCAAAACAGACAGCTGTGTGGAGGAGATCTCTGGACCCACCTGGGTCTCCAGGCA 937
91 AlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySerAsp 110
938 GCCACTGCTTCAGAAACATACCGATGTTCACTGGAGGTGGGCGAGCTCAGAC 997
111 LysLeuGlySerPheProSerIleuAlaValAlaLysIlellellellelleGluPheAsnPro 130
998 AAACCTGGGAGCTTCCATCCCTGGTGTGGCCAAAGATCATCATCATCATCAACCC 1057
131 MetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThraPheSer 150
1058 ATGTACCCCAAGACAAATGACATCGCCCTCATGAAGCTGCAATCTCCACTCACTTCTCA 1117
151 GlyThrValArgProIleCysLeuProPhePheAspGluGluLeuThraProAlaThrPro 170
1118 GGCACAGTCAGGCCCATCTGTCTGCCCTTCTTGTAGAGGAGCTCACTCCAGCCACCCCA 1177
171 LeuTrpIlellelleGlyTrpGlyPheThrLysGlnAsnGlyGlyLysMetSerAspIleLeu 190
1178 CTTCTGGATCATTTGGATGGGGCTTTACGAGCAGAAATGGAGGGAAGATGCTGCACATCTG 1237
191 LeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAspAlaTyrGln 210
1238 CTGCAGGCTCAGTCCAGTCAATGACACACAGCTGCAATGACAGATGCTGCATCTG 1297
211 GlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThrCys 230
1298 GGGGAAGTCACCGAGAAGATGATGTGTGAGGATCTCCGGAAGGGGGGTGTGACACCTGC 1357

S-09-607-745-9 (1-292) x US-09-386-653A-8 (1-1130)

[illegible]

RESULT 5

S-09-386-642-7

Sequence 7, Application US/09386642

```

; Patent No. 6420157
;
; GENERAL INFORMATION:
;
; APPLICANT: Darrow, Andrew
;
; APPLICANT: Qi, Jensen
;
; APPLICANT: Andrade-Gordon, Patricia
;
; TITLE OF INVENTION: Zymogen Activation System
;
; FILE REFERENCE: ORT-1028
;
; CURRENT APPLICATION NUMBER: US/09/386,642
;
; CURRENT FILING DATE: 1999-08-31
;
; NUMBER OF SEQ ID NOS: 60
;
; SOFTWARE: PatentIn Ver. 2.0
;
; SEQ ID NO 7
;
; LENGTH: 1169
;
; TYPE: DNA
;
; ORGANISM: Artificial Sequence
;
; FEATURE:
;
; OTHER INFORMATION: Description of Artificial Sequence: Fusion gene
;
; OTHER INFORMATION: with homo sapien serine protease catalytic domain
;
; US-09-386-642-7

```

Alignment Scores:					
Pred. No.:	5.93e-73	Length:	1169		
Score:	746.00	Matches:	163		
Percent Similarity:	60.37%	Conservative:	35		
Best Local Similarity:	49.70%	Mismatches:	94		
Query Match:	46.62%	Indels:	36		
DB:	4	Gaps:	8		
US-09-607-745-9 (1-232) x US-09-386-642-7 (1-1169)					
QY	1	MetAspSerLySgLySerSerGlnLysSerArgLeuLeuLeuValValSerAen	20		
Db	13	ATGGACGACGAAAGTTGTCGCAGAAATCCCGCTGCTCCTGCTGGTGTGCAAT	72		
QY	21	LeuLeuLeuCysGlnGlyValValSerAspTyrLySAspAspAspValAspAlaA	40		
Db	73	CTACTCTGTGCAGGGTGTGCTCCGACTACAAAGGACGACGACGCGGCGC	132		
QY	41	AlaLeuAlaAlaProPheAspAspAspLysIleValGlyGlyTyrAlaLeuAspVal	60		
Db	133	GCCTCTGCTCCCGCTTGTATGATGATGACAAAGATCGTTGGGGGCTATGCTCTAGAGCC	192		
QY	61	AspSerTrpProTrpGlnValSerIleclnTyrAspLysGlnHisValCysGlyGlySer	80		
Db	193	GGTCAGTGGCCCTGGCAGGTGAGCATCACCTATGAGGCGTCCATGTGTGTGGTGGCTCT	252		
QY	81	IleLeuAspProHisTrpValLeuThrAlaAlaHisCysPheArgLysHisThrAspVal	100		

213 -----ValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThr 229
 673 CCGCATTGTTCACAGAGACATGGGTGTGTGGTATGTGAGAGGGGGCAAGGACGCC 732
 230 CysGlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAsp---GlnTrpHisValVal 248
 733 TGCCAGGGTGACTCTGGGGCCCACTCTCTGCGCTGTGGAGGGTCTCTGGTACCTGACG 792
 249 GlyIleValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyValTyrThrLys 268
 793 GGCATTGTGAGTGGGGAGATGCTGTGGGGCCCGCAACAGGCGCTGGTGTACACTCTG 852
 269 ValSerAlaTyrLeuAsnTrpIleTyr----- 277
 853 GCCTCCAGTATCCCTCTGGATCCAAAGAGGTGACAGAACTCCAGCGCTCGTGTGGTG 912
 278 -----AsnValTrpLysAlaGluLeu----- 284
 913 CCCCACCCAGGAGTCCAGCGCCGACAGCAACTCTGTGGCAGCACTGGCCTTCAGC 972
 285 SerArgHisHisHisHisHisHis 292
 973 TCTAGACATCAACATCACCATCAC 996

RESULT 6

US-09-386-629-2

Sequence 2, Application US/09386629

Patent No. 6426199

GENERAL INFORMATION:

APPLICANT: Darrow, Andrew L.

APPLICANT: Qi, Jensen

APPLICANT: Andrade-Gordon, Patricia

TITLE OF INVENTION: Identification and Characterization of the complementary DNA encoding the novel human serine protease C-E

FILE REFERENCE: ORT-1030

CURRENT APPLICATION NUMBER: US/09/386,629

CURRENT FILING DATE: 1999-08-31

NUMBER OF SEQ ID NOS: 11

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 2

LENGTH: 1166

TYPE: DNA

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: Description of Artificial Sequence: C-E catalytic

OTHER INFORMATION: domain in a zymogen activated construct

US-09-386-629-2

Alignment Scores:

Seq. No.	Score	Length	Matches
2, 35e-72	740.50	1166	164
Conservative	61.70%	39	87
Mismatches	49.85%	39	10
Indels	46.28%	4	
Gaps			

US-09-607-745-9 (1-292) x US-09-386-629-2 (1-1166)

1 MetAspSerLysGlySerSerGlnLysSerArgLeuLeuLeuValValSerAsn 20
 13 ATGGACAGCAAGGTTCTGCGAGAAATCCCGCTCTCTCTGCTGTGTGTGTCAAT 72
 21 LeuLeuLeuCysGlnGlyValValSerAspTyrLysAspAspValAlaAla 40
 73 CTACTCTTGTCCAGGTGTGTCTCCGACTACAGAGCAGCAGCAGTGGAGCGGGC 132
 41 AlaLeuAlaAlaProPheAspAspAspLysIleValGlyTyrAlaLeuAspVal 60
 133 GCTCTGTGCCCCCTTGTATGATGATGACAGATCGTGGGGGCTATCTCTAGAGGAC 192
 61 AspSerTrpProTrpGlnValSerIleGlnTyrAspLysGlnHisValCysGlyGlySer 80

193 AGCGAGTGGCCCTGGATCGTGAGCATCCAGAAAGATGGACCCACCATCGCGAGTTCT 252
 81 IleLeuAspProHisTrpValLeuThrAlaAlaHisCysPheArgLysHisThrAsp--- 99
 253 CTGCTCACCAGCGCGTGGGTGATCACTGCTGCCACTGTTTCAAGGACAACTGAACAAA 312
 100 ValPheAsnTrpLysValArgAlaGlySerAspLysLeuGlySerPheProSerLeuAla 119
 313 CCATACCTGTTCTGTGCTGTGGGGCGCTGGCAGCTGGGAAC---CTGGCTCTCGG 369
 120 ValAlaLysIleIleIle-----IleGluPheAsnProMetTyrPro---LysAspAsn 136
 370 TCCAGAGGTGGGTGTTCCCTGGGTGGAGGCCCACTGCTGTATCTCTGGAGGAGGT 429
 137 -----AspIleAlaLeuMetLysLeuGlnPheProLeuThrPheSerGlyThrVal 153
 430 GCCTGTGCAGACATTCCTGCTGCTCGAGCGCTCCATACAGTTCTCAGACGGGTC 489
 154 ArgProIleCysLeuProPheAspGluGluLeuThrProAlaThrProLeuTrpIle 173
 490 CTGCCCATCTGCTACCTGATGCTCTATCCACCTCCCTCCAAACACCCACCTGCTGATC 549
 174 IleGlyTrpGlyPheThrLysGlnAsnGlyGlyLysMet-----SerAspIleLeu 191
 550 TCAGGCTGGGG---AGCATCCAGATGGAGTTCCTTGGCCCCCACCCTCAGACCTGCGAG 606
 192 GlnAlaSerValGlnValIleAspSerThrArgCysAsnAla-----AspAspAla 208
 607 AAGCTGAAGGTTCTTATCATCATCTCGAAGTCTGAGCCATCTGTACTGGCGGGAGCA 666
 209 TyrGlnGlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAsp 228
 667 GGACAGGACACCATCAGTGGAGCATGTGTGTGCGGCTACTTGGAGGGGAGCGGGAT 726
 229 ThrCysGlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAsp---GlnTrpHisVal 247
 727 GCTGTCTGGCGACTCCGGGGCCCTCATGTGCCAGGTGACGGCCCTGGCTGCTG 786
 248 ValGlyIleValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyValTyrThr 267
 787 GCCGCGCATCATCAGCTGGGGCGAGGCTGTGCCAGCGCAACAGGCGCGGGGTCTACATC 846
 268 LysValSerAlaTyrLeuAsnTrpIleTyrAsnValTrpLys----- 281
 847 AGCTCTCTGCGCACCGCTCTCTGGGTGGAGAGATCTGTGCAAGGGGTGCGACTCCCGGG 906
 282 -----AlaGlu 283
 907 CGCGCTCAGGGGGTGGGGCTCTCAGGGCACCGAGGCTCTGGGGCGCGCGCGCGC 966

RESULT 7

US-09-386-642-10

Sequence 10, Application US/09386642

Patent No. 6420157

GENERAL INFORMATION:

APPLICANT: Darrow, Andrew

APPLICANT: Qi, Jensen

APPLICANT: Andrade-Gordon, Patricia

TITLE OF INVENTION: Zymogen Activation System

FILE REFERENCE: ORT-1028

CURRENT APPLICATION NUMBER: US/09/386,642

CURRENT FILING DATE: 1999-08-31

NUMBER OF SEQ ID NOS: 60

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 10

LENGTH: 1052

TYPE: DNA

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: Description of Artificial Sequence: Fusion gene
OTHER INFORMATION: with homo sapien serine protease catalytic domain

us-09-386-642-10

Alignment Scores:

6.43e-65 Length: 1052
Score: 672.50 Matches: 148
Percent Similarity: 61.31% Conservative: 39
Best Local Similarity: 48.52% Mismatches: 89
Query Match: 42.03% Indels: 29
Gaps: 11

US-09-607-745-9 (1-292) x US-09-386-642-10 (1-1052)

```
Y 1 MetAspSerLysGlySerGlnLysSerArgLeuLeuLeuLeuValValSerAsn 20
b 13 ATGGACAGCAAAAGTTCTGCGCAGAAATCCCGCTCTCTCTGCTGCTGTGTCACAAAT 72
Y 21 LeuLeuLeuCysGlnGlyValValSerAspTyrLysAspAspValAspAla 40
b 73 CTACTCTTGCCAGGGTGTGGTCCGACTACAGGACGACGACGTCGACGCGGCC 132
Y 41 AlaLeuAlaAlaProPheAspAspAspLysIleValGlyGlyTyr--AlaLeuAsp 59
b 133 GCTCTTGCTGCCCCCTTTGATGATGACAGATGTTGGGGCTACAACTGTCTAGAA 192
Y 60 ValAspSerTrpProTrpGlnValSerIleGlnTyrAspLysGlnHisValCysGly 79
b 193 AAGCACTCCAGCCCTGGCAGGCGAGCCCTGTCAGAGACGCGCTACTCTGTGGGGCG 252
Y 80 SerIleLeuAspProHisTrpValLeuThrAlaAlaHisCysPheArgLys----- 96
b 253 AGCTCATCGCCCGCAGTGGCTTCCTGAGAGAGCCCACTGCTCAAGCCCGCGCTACATA 312
Y 97 ---HisThrAspValPheAsnTrpLysValArgAlaGlySerAspLys----- 111
b 313 GTTCACTGGGCGCAGCACACCTCCAGAGGAGGAGGCTGTGAGCAGACCCGCGACGCC 372
Y 112 LeuGlySerPheProSerLeuAlaValAlaLysIleIleIleGlnPheAsnProMet 131
b 373 ACTGATCTCTCCCGCCACCCCGGC-----TTCAACACACAGC 408
Y 132 TyrPro---LysAsp-----AsnAspIleAlaLeuMetLysLeuGlnPheProLeuThr 148
b 409 CTCGCCCAACAGACACCCGCAATGACATCTGCTGTGAGATGCATCGCAGTCTCC 468
Y 149 PheSerGlyThrValArgProIleCysLeuProPheAspGluLeuThrProAla 168
b 469 ATCACTGGGCTGTGGACCCCTCACTCC-----TCAGCTGTGTCACTGTGGC 522
Y 169 ThrProLeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyLysMetSerAsp 188
b 523 ACCAGTGCTCTATTCCGGCTGGGCGACAGCTCCAGCCCGCCAGTTACGCTGCTCAC 582
Y 189 IleLeuLeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAla 208
b 583 ACCTGCGATGGCCCAACATCACCATTGATGACACCAAGT-----GAGACGCC 636
Y 209 TyrGlnGlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyValAsp 228
b 637 TACCCCGGCAACATCACAGACACCATGTTGTGCCAGGTGCAGGAAGGGGCGAGGAC 696
Y 229 ThrCysGlnGlyAspSerGlyLysProLeuMetTyrGlnSerAspGlnTrpHisVal 248
b 697 TCTTGCGAGGGTACTCCGGGGGCCCTCTGTGTGTATACCACTCT-----CTTCAA 747
Y 249 GlyIleValSerTrpGlyTyrGly---CysGlyGlyProSerThrProGlyValTrpThr 267
b 748 GGCATTATCTCTGGGGCCAGGATCCGTGTGGATCACCCGAAAGCCCTGTGTACACG 807
Y 268 LysValSerAlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeuSerArgHis 287
b 808 AAAGTCGCAATATGTGGACTGGATCCAGGAGCAGCATGAGACAAAT---TCTAGACAT 864
```

QY 288 HisHisHisHisHis 292
Db 865 CACCATCACCATCAC 879

RESULT 8

US-09-386-642-9
; Sequence 9, Application US/09386642
; Patent No. 6420157
; GENERAL INFORMATION:
; APPLICANT: Darrow, Andrew
; APPLICANT: Qi, Jensen
; APPLICANT: Andrade-Gordon, Patricia
; TITLE OF INVENTION: Zymogen Activation System
; FILE REFERENCE: ORT-1028
; CURRENT APPLICATION NUMBER: US/09/386,642
; CURRENT FILING DATE: 1999-08-31
; NUMBER OF SEQ ID NOS: 60
; SOFTWARE: PatentIn ver. 2.0
; SEQ ID NO 9
; LENGTH: 1049
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Fusion gene
; OTHER INFORMATION: with homo sapien serine protease catalytic domain
US-09-386-642-9

Alignment Scores:

Pred. No.: 6,92e-62 Length: 1049
Score: 645.00 Matches: 146
Percent Similarity: 60.33% Conservative: 38
Best Local Similarity: 47.87% Mismatches: 91
Query Match: 40.31% Indels: 30
DB: 4 Gaps: 11

US-09-607-745-9 (1-292) x US-09-386-642-9 (1-1049)

```
QY 1 MetAspSerLysGlySerSerGlnLysSerArgLeuLeuLeuValValSerAsn 20
Db 13 ATGGACAGCAAAAGTTCTGCGCAGAAATCCCGCTCTCTCTGCTGTGTCACAAAT 72
QY 21 LeuLeuLeuCysGlnGlyValValSerAspTyrLysAspAspValAspAla 40
Db 73 CTACTCTTGCCAGGGTGTGGTCTCCGACTACAGGACGACGACGTCGACGCGGCC 132
QY 41 AlaLeuAlaAlaProPheAspAspAspLysIleValGlyGlyTyr---AlaLeuAsp 59
Db 133 GCTCTTGCTGCCCCCTTTGATGATGACAGATCTGTTGGGGCTACAACTGTCTAGAA 192
QY 60 ValAspSerTrpProTrpGlnValSerIleGlnTyrAspLysGlnHisValCysGly 79
Db 193 CCCCATTCGCGACCTTGGCAGGCGCCTTGTCCAGGCCAGCACTACTCTGTGGCGGT 252
QY 80 SerIleLeuAspProHisTrpValLeuThrAlaAlaHisCysPheArg---LysHisThr 98
Db 253 GTCCTTGTAGTGGCACTGGGTCTCTTACAGTCCCACTGTAAAAAACCGAAATACACA 312
QY 99 AspValPheAsnTrpLysValArgAlaGlySerAspLysLeu-----GlySer 114
Db 313 -----GTACGCTGGGAGACGACGCTACAGATAAAGATGGCCCA 354
QY 115 PheProSerLeuAlaValAlaLysIleIleIleGluPheAsnProMetTyr----- 132
Db 355 GAGCAAGAAATACCTGTGGTTCAGTCCATCCCA-----CACCCCTGTACACAGC 405
QY 133 -----ProLysAspAspAspIleAlaLeuMetLysLeuGlnPheProLeuThr 148
Db 406 AGCGATGTGGAGGACCAACACCATGATCTGATGCTTCTTCACTGCGGTGACGAGCATCC 465
QY 149 PheSerGlyThrValArgProIleCysLeuProPheAspGluLeuThrProAla 168
Db 466 CTGGGGTCCAAAGTGAAGCCCATCAGCCTTGCA-----GATCATTCACCCCGCCTGGC 519
```

Db Db

187 AGCGTCGGCTGTGGGATCCCAAGTATGCCGAGTGAGCCTGCCTACGACCACCGCTGGGCA 246

QY LeuThrAlaIaHisCysPheArgLysHisThrAspValPheAsn-----TrpLys 104

Db 247 CTCACGGGGCCCACTGCTTTGAACCTATAGTAGACCTTAGTGTATCCCTCCGGGTGGATG 306

QY ValArgAlaGlySerAspLysLeuGlySerPheProSerLeuAlaValAlalys----- 122

Db 307 GTCCAGATTGGC-----CAGCTGACTTCCATGCCATCCCTCTGGAGCCTGCAGGCGCTAC 360

QY 123 -----IllellellelleGluPheAsnProMetTyrProLysAspAsn--- 136

Db 361 TACAACCGTTACTTCGTATCGAATATATCTAGGACCTCGCTACCTGGGGAATCACCC 420

QY 137 ---AspilleAlaLeuMetLysLeuGlnPheProLeuThrPheSerGlyThrValArgPro 155

Db 421-TATGACATTGCTTGGTGAAGCTGTTCGACCTGTCCACCTACACTAAACACATCCAGCCC 480

QY 156 lleCysLeuProPhePheAspGluLeuThrProAlaThrProLeuTrpIleleGly 175

Db 481 ATCTGCTCCAGGCTCCACATTTGAGTTTGAGAACCGGACAGACTGCTGGGTGACTGGC 540

QY 176 TrpGlyPheThrLysGlnAsnGlyGlyLysMetSerAspIleLeuLeuGln---AlaSer 194

Db 541 TGGGGGTACATAAAGAGGATGAGCGCACTGCCATCTCCCCACACCCTCCAGGAAGTTTCAG 600

QY 195 ValGlnValIleAspSerThrArgCysAsp-----AlaAspAspAlaTyrrGlnGly 211

Db 601 GTGCGCATCATAAACACTCATGTGCAACCAACCTTCTCCTCAGTACAGTTTCCGCAAG 660

QY 212 GluValThrGluLysMetMetCysAlaGlyIleProGluGlyValAspThrCysGln 231

Db 661 GACATCTTTGGAGACATGTTTGTGCTGGCAATGCCCAAGCGGGAAGGATGCTGCTTC 720

QY 232 GlyAspSerGlygProLeu---MetTyrGlnSerAspGlnTrpHisValValGlyIle 250

Db 721 GGTAAGTTCAGGTGGACCCCTTGGCCCTGTAAACAAGAANTGCACTGGTATCAGATTGGAGTC 780

QY 251 ValSerTrpGlyTyrrGlyCysglyGlyProSerThrProGlyValTyrrThrLysValSer 270

Db 781 GTGAGCTGGGAGATGGGCTGTGGTGGCGCCAATCGGCCCGGTGTCTACACCAATATCAGC 840

QY 271 AlaTyrrLeuAsnTrpIleTyrrAsnVal----- 279

Db 841 CACCACCTTTGAGTGGATCCAGAACTGATGCCCGCAGAGTGGCATGTCCCGACGACGCCCC 900

QY 280 ----TrpLysAlaGluLeuSerArgHisHisHisHisHis 292

Db 901 TCCTCG-----TCTAGACATCACCATCACCATCACC 930

OTHER INFORMATION: Description of Artificial Sequence: Nucleic acid sequence of human protease F in CPEK2 zymogen

28 ValSerAspTyrIysAspAspAspValAspAlaIaIaLeuAlaIaProPheAsp 47
73 GTCCCCGACTACAGACACACAC-----GACGGCGGCTCTGTGCGCCGCTTGTAT 126
48 AspAspLysIleValGlyGlyTyrAlaLeuAspValAspSerTtpProTtpGlnVal 67
127 GATGATGACAGATCGTTGGGGGCTATGCTCTAGACACTCGGGCGCTTGGCGTGGCAGGGG 186
68 SerIeclnTyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTtpVal 87
:::~::~:

532 GGCTGGGCTCTGCTGGCG-----AACGGCAGAAATGCTACCGTGTGCGTGGTGAAC 585
195 ValGlnValIleAspSerThrArgCysAsnAla-----AspAspAlaThrGlnGlyGlu 212
586 GTGTCGGTGGTGTCTGAGAGGCTCTGAGTAAGCTTATGACCCGCTGTACCAC----- 639
213 ValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThrCysGlnGly 232
640 -----CCACGATGTTCTGCGCGCGGGAGGCGACACAGAGACTCTCTCAACGGT 693
233 AspSerGlyGlyProLeuMetThrGlnSerAspGlnTrpHisValValGlyIleValSer 252
694 GACTCTGGGGGGCCCTGTGCTGCAACGGG-----TACTGTGAGGGCCTGTGTCT 744
253 TrpGlyThrGly---CysGlyGlyProSerThrProGlyValThrThrLysValSerAla 271
745 TTCGGAAAGCCCGTGTGCCAAGTTGGCGTCCAGGTGTCTACACCACTCTGCAAA 804
272 TyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeuSerArgHisHisHis 291
805 TTCACCTGAGTGGTAGAGAAACCGTCCAGGCCAGT---TCTAGACATCACCATCCAT 861
292 His 292
862 CAC 864

RESULT 12
US-09-370-838-80
Sequence 80, Application US/09370838
Patent No. 644425
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Lodes, Michael J.
APPLICANT: Mohanath, Roadoh
APPLICANT: Secrist, Heather
TITLE OF INVENTION: COMPOUNDS FOR THERAPY AND DIAGNOSIS OF
FILE REFERENCE: 210121.475C1
CURRENT APPLICATION NUMBER: US/09/370,838
EARLIER FILING DATE: 1999-08-09
EARLIER FILING DATE: 1999-04-02
NUMBER OF SEQ ID NOS: 289
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 80
LENGTH: 1460
TYPE: DNA
ORGANISM: Homo sapien
-09-370-838-80
ignment Scores:
ed. No.: 1,318-48 Length: 1460
ore: 526.50 Matches: 97
cent Similarity: 62.93% Conservative: 49
st Local Similarity: 41.81% Mismatches: 79
ery Match: 32.91% Indels: 7
4
Gaps: 4
-09-607-745-9 (1-292) x US-09-370-838-80 (1-1460)
49 AspAspLysIleValGlyGlyTyrAlaLeuAspValAspSerTrpTrpTrpValSer 68
601 GAGCAGAGATCCTTGGAGGCACTGAGCTGAGGAGGAGGAGTGGCGGTGGCACTGAGT 660
69 IleGlnTyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeu 88
661 CTGCGCTCAATAATATCCACACCACTGTGGAGGCGCTGATCAATCAATGATGATCTG 720
89 ThrAlaAlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGly 108
721 ACAGAGCTCACTGCTTCTGAGAGCAACTTAATCTCTGCTGACTGATTCACCACTGCT 780
109 SerAspLysLeuGlySerPheProSerLeu-----AlaValAlaLysIleIleIle 126

Db 781 ATTTCACA-----ACATTTCTAACTAAAGATGAGAGTAAAGAAATATTTAATTCAT 834
QY 127 GluPheAsnProMetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPhePro 146
Db 835 AACAAATTAAATCTGCAACTCATGAAATGACATGCTGCTGTGAGACTTGAGAACAGT 894
QY 147 LeuThrPheSerGlyThrValArgProLysCysLeuProPhePheAspGluGluLeuThr 166
Db 895 GTACACTTTTACCAAGATATCCATAGTGTGTCTCCACAGCTCTACCCAGAATAATCCA 954
QY 167 ProAlaThrProLeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyGlyLysMet 186
Db 955 CTGGCTCTACTGCTTATGTAACAGATGGGGCGTCTCAAGATATGCTGCCACACAGTT 1014
QY 187 SerAspIleLeuLeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAsp 206
Db 1015 CCAGAG---CTAAGCCAGGACAGCTCAGAAATAAAGTAATGATGATGATGATGATGAT 1071
QY 207 AspAlaThrGlnGlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyGly 226
Db 1072 CATAGTTAATGAGGACCATCTTGTCTGGAATGCTGTGTCTGCTGAGTACCTCAAGGTGA 1131
QY 227 ValAspThrCysGlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGln----- 244
Db 1132 GTGGACCATGTCAGGGTGAATCTGTGGTGGCCACTAGTACAGAGAGACTCAAGGGGCTT 1191
QY 245 TrpHisValValGlyIleValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGly 264
Db 1192 TGGTTTATTTGGGATAGTAAGCTGGGAGATCAGTGTGGCTCCGGATAGCCAGGA 1251
QY 265 ValTyrThrLysValSerAlaTyrIleuAsnTrpIle 276
Db 1252 GTGTATCTCGAGTGACAGCCTACCTTCACTGGATT 1287

RESULT 13
US-08-508-448C-15
Sequence 15, Application US/08508448C
Patent No. 5804410
GENERAL INFORMATION:
APPLICANT: Kazuyoshi YAMAOKA et al.
TITLE OF INVENTION: NUCLEIC ACID SEQUENCE ENCODING
TITLE OF INVENTION: TRYPsin-LIKE ENZYME AND PROCESS FOR PRODUCING THE ENZYME
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: Wenderoth, Lind & Ponack
STREET: 805 Fifteenth Street, N.W., #700
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
COMPUTER: IBM Compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/508,448C
FILING DATE: July 28, 1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Warren M. Cheek, Jr.
REGISTRATION NUMBER: 33,367
REFERENCE/DOCKET NUMBER:
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-371-8850
TELEFAX:
TELEX:
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

M protein - nucleic search, using frame_plus_p2n model

run on: November 23, 2003, 07:54:55 ; Search time 314.091 Seconds
(without alignments)
3038.343 Million cell updates/sec

Result score: 1600
Sequence: 1 MDSGSSQKSRLLLLLVSN.....LNWYNVWKAELSRHHHHH 292

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2169961 seqs, 1634102185 residues

Total number of hits satisfying chosen parameters: 4339922

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

MODEL=frame+ p2n.model -DEV=xlh
DB=Published Applications NA -QFMT=fastap -SUPFI=xnpb -MINMATCH=0.1
-COOCL=0 -LOOPEXT=0 -UNIS=bits -START=1 -END=1 -MATRIX=blonum62
-TRANS=human40.cdi -LIST=45 -DOCLALIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0
-MAXLEN=200000000 -USER=US09607745 @CGN 1 1 424 @runat 21112003 144347 22270
-ICPU=3 -NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOC=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-XGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA:*

1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
15: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
16: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
17: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	1266	79.1	2038 12	US-10-180-719-18 Sequence 18, Appl

2	1266	79.1	2063	11	US-09-888-257A-2	Sequence 2, Appl
3	1266	79.1	2063	11	US-09-888-257A-2	Sequence 274, App
4	1266	79.1	2063	12	US-10-015-387A-274	Sequence 274, App
5	1266	79.1	2063	12	US-10-063-735-111	Sequence 111, App
6	1266	79.1	2063	12	US-10-006-130A-274	Sequence 274, App
7	1266	79.1	2063	12	US-10-199-672-329	Sequence 329, App
8	1266	79.1	2063	12	US-10-006-172A-274	Sequence 274, App
9	1266	79.1	2063	12	US-10-187-749-329	Sequence 329, App
10	1266	79.1	2063	12	US-10-194-457-329	Sequence 329, App
11	1266	79.1	2063	12	US-10-184-642-329	Sequence 329, App
12	1266	79.1	2063	12	US-10-196-747-329	Sequence 329, App
13	1266	79.1	2063	12	US-10-015-392A-274	Sequence 274, App
14	1266	79.1	2063	12	US-10-017-253A-274	Sequence 274, App
15	1266	79.1	2063	12	US-10-173-689-329	Sequence 329, App
16	1266	79.1	2063	12	US-10-173-690-329	Sequence 329, App
17	1266	79.1	2063	12	US-10-173-691-329	Sequence 329, App
18	1266	79.1	2063	12	US-10-173-692-329	Sequence 329, App
19	1266	79.1	2063	12	US-10-173-694-329	Sequence 329, App
20	1266	79.1	2063	12	US-10-173-698-329	Sequence 329, App
21	1266	79.1	2063	12	US-10-173-699-329	Sequence 329, App
22	1266	79.1	2063	12	US-10-173-707-329	Sequence 329, App
23	1266	79.1	2063	12	US-10-174-569-329	Sequence 329, App
24	1266	79.1	2063	12	US-10-174-583-329	Sequence 329, App
25	1266	79.1	2063	12	US-10-174-587-329	Sequence 329, App
26	1266	79.1	2063	12	US-10-174-589-329	Sequence 329, App
27	1266	79.1	2063	12	US-10-174-591-329	Sequence 329, App
28	1266	79.1	2063	12	US-10-175-736-329	Sequence 329, App
29	1266	79.1	2063	12	US-10-175-742-329	Sequence 329, App
30	1266	79.1	2063	12	US-10-175-744-329	Sequence 329, App
31	1266	79.1	2063	12	US-10-175-745-329	Sequence 329, App
32	1266	79.1	2063	12	US-10-175-748-329	Sequence 329, App
33	1266	79.1	2063	12	US-10-175-751-329	Sequence 329, App
34	1266	79.1	2063	12	US-10-175-754-329	Sequence 329, App
35	1266	79.1	2063	12	US-10-176-480-329	Sequence 329, App
36	1266	79.1	2063	12	US-10-176-489-329	Sequence 329, App
37	1266	79.1	2063	12	US-10-176-754-329	Sequence 329, App
38	1266	79.1	2063	12	US-10-176-755-329	Sequence 329, App
39	1266	79.1	2063	12	US-10-176-920-329	Sequence 329, App
40	1266	79.1	2063	12	US-10-176-922-329	Sequence 329, App
41	1266	79.1	2063	12	US-10-176-922-329	Sequence 329, App
42	1266	79.1	2063	12	US-10-176-924-329	Sequence 329, App
43	1266	79.1	2063	12	US-10-176-984-329	Sequence 329, App
44	1266	79.1	2063	12	US-10-179-508-329	Sequence 329, App
45	1266	79.1	2063	12	US-10-179-512-329	Sequence 329, App

ALIGNMENTS

RESULT 1

US-10-180-719-18
; Sequence 18, Application US/10180719
; Publication NO. US20030166246A1
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; Hillman, Jennifer L.
; Yue, Henry
; Guegler, Karl J.
; Corley, Neil C.
; Tang, Tom Y.
; Shah, Purvi

TITLE OF INVENTION: HUMAN PROTEASE MOLECULES
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESS: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Dr.
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM compatible
OPERATING SYSTEM: DOS

-09-607-745-9 (1-292) X US-09-888-257A-2 (1-2063)

51 LysIleValGlyGlyTyrAlaLeuAspValAspSerTrpProTrpGlnValSerIleGln 70
: : : : :
804 CGTGTGGTGGTGGGAGGAGGCTCTGTGGATTCTTGGCCCTGGCAGGTCCAGTCACG 863

71 TyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThrAla 90
: : : : :
864 TACGCAAAACACACAGCTCTGTGGAGGAGCATCTCGACCCCATCTGGGTCTCCAGGCA 923

91 AlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySerAsp 110
924 GCCCATCTGCTCAGGAACAATACCGATGTTCACTGGAAGGTGGGCGACGCTCAGAC 983

111 LysLeuGlySerPheProSerIleuAlaValAlaLysIleIleIleGluPheAsnPro 130
984 AAACCTGGGCGAGCTTCCCATCCCTGGCTGTGGCGCAAGATCATCATCTGAATCAACCC 1043

131 MetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPheSer 150
1044 ATGTACCCCAAGACAATGACATCCCCCTCATGAAGCTCAGTTCCTCCACTCTTCTCA 1103

151 GlyThrValArgProIleCysLeuProPhePheAspGluGluLeuThrProAlaThrPro 170
1104 GGCACAGTCAGGCCCATCTGTCTGCCCTTCTTTGATGAGGAGCTCACTCCAGCCACCCCA 1163

171 LeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyGlyLysMetSerAspIleLeu 190
1164 CTCTGGATCATGTGATGGGCTTTACGAGCAGAGATGGAGGAAGATGCTGACATCTG 1223

191 LeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAlaTrpGln 210
1224 CTGAGGGCTGATCCAGTCAATGACAGCACCGGTGCAATGCGAGAGTACCGTACCCAG 1283

211 GlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThrCys 230
1284 GGGGAAGTCACCGAAGATGATGTGTGCGAGCATCCCGGAAGGGGGTGTGACACCTGC 1343

231 GlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGlyIle 250
1344 CAGGTGACAGTGTGGGCCCTGATGTACCAATCTGACCACTGGCATGTGTGGGCGATC 1403

251 ValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyValTyrThrLysValSer 270
1404 GTTAGCTGGGGCTATGGTGGGGGGCCCGAGCACCCCGAGGATATACACCAAGGTCTCA 1463

271 AlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu 284
1464 GCCTATCTCAACTGGATCTACATGTCTGGAAGGTGAGCTG 1505

SULT 3

-09-946-374-274

Sequence 274, Application US/09946374

Publication No. US20030073129A1

GENERAL INFORMATION:

APPLICANT: Baker, Kevin P.

APPLICANT: Botstein, David

APPLICANT: Desnoyers, Luc

APPLICANT: Eaton, Dan L.

APPLICANT: Ferrara, Napoleone

APPLICANT: Fong, Sherman

APPLICANT: Gao, Wei-Qiang

APPLICANT: Goddard, Audrey

APPLICANT: Godowski, Paul J.

APPLICANT: Grimaldi, Christopher J.

APPLICANT: Gurney, Austin L.

APPLICANT: Hillan, Kenneth J.

APPLICANT: Pan, James

APPLICANT: Paoni, Nicholas F.

APPLICANT: Roy, Margaret Ann

APPLICANT: Smith, Victoria

APPLICANT: Stewart, Timothy A.

APPLICANT: Tumas, Daniel

APPLICANT: Watanabe, Colin K.
APPLICANT: Williams, P. Mickey
APPLICANT: Wood, William I.
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
TITLE OF INVENTION: Acids Encoding the Same
FILE REFERENCE: P2830FIC1
CURRENT APPLICATION NUMBER: US/09/946,374
CURRENT FILING DATE: 2001-09-04
PRIOR APPLICATION NUMBER: 60/098716
PRIOR FILING DATE: 1998-09-01
PRIOR APPLICATION NUMBER: 60/098723
PRIOR FILING DATE: 1998-09-01
PRIOR APPLICATION NUMBER: 60/098749
PRIOR FILING DATE: 1998-09-01
PRIOR APPLICATION NUMBER: 60/098750
PRIOR FILING DATE: 1998-09-01
PRIOR APPLICATION NUMBER: 60/098803
PRIOR FILING DATE: 1998-09-02
PRIOR APPLICATION NUMBER: 60/098821
PRIOR FILING DATE: 1998-09-02
PRIOR APPLICATION NUMBER: 60/098843
PRIOR FILING DATE: 1998-09-02
PRIOR APPLICATION NUMBER: 60/099536
PRIOR FILING DATE: 1998-09-03
PRIOR APPLICATION NUMBER: 60/099596
PRIOR FILING DATE: 1998-09-03
PRIOR APPLICATION NUMBER: 60/099598
PRIOR FILING DATE: 1998-09-03
PRIOR APPLICATION NUMBER: 60/099602
PRIOR FILING DATE: 1998-09-03
PRIOR APPLICATION NUMBER: 60/099642
PRIOR FILING DATE: 1998-09-03
PRIOR APPLICATION NUMBER: 60/099741
PRIOR FILING DATE: 1998-09-10
PRIOR APPLICATION NUMBER: 60/099754
PRIOR FILING DATE: 1998-09-10
PRIOR APPLICATION NUMBER: 60/099763
PRIOR FILING DATE: 1998-09-10
PRIOR APPLICATION NUMBER: 60/099792
PRIOR FILING DATE: 1998-09-10
PRIOR APPLICATION NUMBER: 60/099808
PRIOR FILING DATE: 1998-09-10
PRIOR APPLICATION NUMBER: 60/099812
PRIOR FILING DATE: 1998-09-10
PRIOR APPLICATION NUMBER: 60/099815
PRIOR FILING DATE: 1998-09-10
PRIOR APPLICATION NUMBER: 60/099816
PRIOR FILING DATE: 1998-09-10
PRIOR APPLICATION NUMBER: 60/100385
PRIOR FILING DATE: 1998-09-15
PRIOR APPLICATION NUMBER: 60/100388
PRIOR FILING DATE: 1998-09-15
PRIOR APPLICATION NUMBER: 60/100390
PRIOR FILING DATE: 1998-09-15
PRIOR APPLICATION NUMBER: 60/100584
PRIOR FILING DATE: 1998-09-16
PRIOR APPLICATION NUMBER: 60/100627
PRIOR FILING DATE: 1998-09-16
PRIOR APPLICATION NUMBER: 60/100661
PRIOR FILING DATE: 1998-09-16
PRIOR APPLICATION NUMBER: 60/100662
PRIOR FILING DATE: 1998-09-16
PRIOR APPLICATION NUMBER: 60/100664
PRIOR FILING DATE: 1998-09-16
PRIOR APPLICATION NUMBER: 60/100683
PRIOR FILING DATE: 1998-09-17
PRIOR APPLICATION NUMBER: 60/100684
PRIOR FILING DATE: 1998-09-17
PRIOR APPLICATION NUMBER: 60/100710
PRIOR FILING DATE: 1998-09-17
PRIOR APPLICATION NUMBER: 60/100711
PRIOR FILING DATE: 1998-09-17
PRIOR APPLICATION NUMBER: 60/100848

PRIOR FILING DATE: 1998-09-18
PRIOR APPLICATION NUMBER: 60/100849
PRIOR FILING DATE: 1998-09-18
PRIOR APPLICATION NUMBER: 60/100919
PRIOR FILING DATE: 1998-09-17
PRIOR APPLICATION NUMBER: 60/100930
PRIOR FILING DATE: 1998-09-17
PRIOR APPLICATION NUMBER: 60/101014
PRIOR FILING DATE: 1998-09-18
PRIOR APPLICATION NUMBER: 60/101068
PRIOR FILING DATE: 1998-09-18
PRIOR APPLICATION NUMBER: 60/101071
PRIOR FILING DATE: 1998-09-18
PRIOR APPLICATION NUMBER: 60/101279
PRIOR FILING DATE: 1998-09-22
PRIOR APPLICATION NUMBER: 60/101471
PRIOR FILING DATE: 1998-09-23
PRIOR APPLICATION NUMBER: 60/101472
PRIOR FILING DATE: 1998-09-23
PRIOR APPLICATION NUMBER: 60/101474
PRIOR FILING DATE: 1998-09-23
PRIOR APPLICATION NUMBER: 60/101475
PRIOR FILING DATE: 1998-09-23
PRIOR APPLICATION NUMBER: 60/101476
PRIOR FILING DATE: 1998-09-23
PRIOR APPLICATION NUMBER: 60/101477
PRIOR FILING DATE: 1998-09-23
PRIOR APPLICATION NUMBER: 60/101479
PRIOR FILING DATE: 1998-09-23
PRIOR APPLICATION NUMBER: 60/101738
PRIOR FILING DATE: 1998-09-24
PRIOR APPLICATION NUMBER: 60/101741
PRIOR FILING DATE: 1998-09-24
PRIOR APPLICATION NUMBER: 60/101743
PRIOR FILING DATE: 1998-09-24
PRIOR APPLICATION NUMBER: 60/101915
PRIOR FILING DATE: 1998-09-24
PRIOR APPLICATION NUMBER: 60/101916
PRIOR FILING DATE: 1998-09-24
PRIOR APPLICATION NUMBER: 60/102207
PRIOR FILING DATE: 1998-09-29
PRIOR APPLICATION NUMBER: 60/102240
PRIOR FILING DATE: 1998-09-29
PRIOR APPLICATION NUMBER: 60/102307
PRIOR FILING DATE: 1998-09-29
PRIOR APPLICATION NUMBER: 60/102330
PRIOR FILING DATE: 1998-09-29
PRIOR APPLICATION NUMBER: 60/102331
PRIOR FILING DATE: 1998-09-29
PRIOR APPLICATION NUMBER: 60/102484
PRIOR FILING DATE: 1998-09-30
PRIOR APPLICATION NUMBER: 60/102487
PRIOR FILING DATE: 1998-09-30
PRIOR APPLICATION NUMBER: 60/102570
PRIOR FILING DATE: 1998-09-30
PRIOR APPLICATION NUMBER: 60/102571
PRIOR FILING DATE: 1998-09-30
PRIOR APPLICATION NUMBER: 60/102684
PRIOR FILING DATE: 1998-10-01
PRIOR APPLICATION NUMBER: 60/102687
PRIOR FILING DATE: 1998-10-01
PRIOR APPLICATION NUMBER: 60/102965
PRIOR FILING DATE: 1998-10-02
PRIOR APPLICATION NUMBER: 60/103258
PRIOR FILING DATE: 1998-10-06
PRIOR APPLICATION NUMBER: 60/103314
PRIOR FILING DATE: 1998-10-07
PRIOR APPLICATION NUMBER: 60/103315
PRIOR FILING DATE: 1998-10-07
PRIOR APPLICATION NUMBER: 60/103328
PRIOR FILING DATE: 1998-10-07
PRIOR APPLICATION NUMBER: 60/103395
PRIOR FILING DATE: 1998-10-07

/ PRIOR APPLICATION NUMBER: 60/103396
/ PRIOR FILING DATE: 1998-10-07
/ PRIOR APPLICATION NUMBER: 60/103401
/ PRIOR FILING DATE: 1998-10-07
/ PRIOR APPLICATION NUMBER: 60/103449
/ PRIOR FILING DATE: 1998-10-06
/ PRIOR APPLICATION NUMBER: 60/103633
/ PRIOR FILING DATE: 1998-10-08
/ PRIOR APPLICATION NUMBER: 60/103678
/ PRIOR FILING DATE: 1998-10-08
/ PRIOR APPLICATION NUMBER: 60/103679
/ PRIOR FILING DATE: 1998-10-08
/ PRIOR APPLICATION NUMBER: 60/103711
/ PRIOR FILING DATE: 1998-10-08
/ PRIOR APPLICATION NUMBER: 60/104257
/ PRIOR FILING DATE: 1998-10-14
/ PRIOR APPLICATION NUMBER: 60/104987
/ PRIOR FILING DATE: 1998-10-20
/ PRIOR APPLICATION NUMBER: 60/105000
/ PRIOR FILING DATE: 1998-10-20
/ PRIOR APPLICATION NUMBER: 60/105002
/ PRIOR FILING DATE: 1998-10-20
/ PRIOR APPLICATION NUMBER: 60/105104
/ PRIOR FILING DATE: 1998-10-21
/ PRIOR APPLICATION NUMBER: 60/105169
/ PRIOR FILING DATE: 1998-10-22
/ PRIOR APPLICATION NUMBER: 60/105266
/ PRIOR FILING DATE: 1998-10-22
/ PRIOR APPLICATION NUMBER: 60/105693
/ PRIOR FILING DATE: 1998-10-26
/ PRIOR APPLICATION NUMBER: 60/105694
/ PRIOR FILING DATE: 1998-10-26
/ PRIOR APPLICATION NUMBER: 60/105807

Alignment Scores:

Pred. No.: 1,6e-152 Length: 2063
Score: 1266.00 Matches: 228
Percent Similarity: 98.29% Conservative: 2
Best Local Similarity: 97.44% Mismatches: 4
Query Match: 79.12% Indels: 0
DB: 11 Gaps: 0

US-09-607-745-9 (1-292) x US-09-946-374-274 (1-2063)

QY	51	LysileValGlyGlyTyrAlaLeuAspValAspSerTrpProTrpGlnValSerIleGln	70
Db	804	CGTGTGGTGGTGGGAGGAGGCGCTCTGTGATTCTTGCCCTTGGCAGGTGAGCATCCAG	863
QY	71	TyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThrAla	90
Db	864	TACGACAAACAGCAGCTGTGTGGAGGGAGCATCTGTGGACCCCACTGGGTCTCTCACGGCA	923
QY	91	AlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySerAsp	110
Db	924	GCCCACTCTTCAGAAACATACCGATGTGTCACTGGAAGGTGGGGAGGCTCAGAC	983
QY	111	LysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleIleGluPheAsnPro	130
Db	984	AAACTGGGACGTTCCCATCCTCGCTGTGGCCAGATCATCATCATTAATTCAACCCC	1043
QY	131	MetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPheSer	150
Db	1044	ATGTACCCCAAGACAATGACATCGCCCTCATGAAGCTGAGTTCACACTCACATTCTCA	1103
QY	151	GlyThrValArgProIleCysLeuProPhePheAspGluLeuLeuThrProAlaThrPro	170
Db	1104	GGCAGTCAGGCCCATCTGTCTCCCTTCTTGTATGAGGAGTCACTCCAGCCACCCCA	1163
QY	171	LeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyGlyLysMetSerAspIleLeu	190
Db	1164	CTCTGGATCAATTGGATGGGGCTTTACGACGAGATGGAGGAAGATGTTCTGACATCTG	1223
QY	191	LeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAlaThrGln	210

1224 CTGACGGGTCAGTCCAGGTCATTGACAGACACAGGTCGAATCCAGACGATGCGTACCAG 1283
 211 GYCluValThrGluLysMetMetCysAlaGlyIleProGluGlyValAspThrCys 230
 1284 GGGGAAGTCACCGAGAAGATGATGTGTGACAGGCATCCCGAAGGGGTGTGGACACTGC 1343
 231 GlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspClnTrpHisValValGlyIle 250
 1344 CAGGTCACAGTGGTGGGCCCCCTGTATGATACCAATCTGACAGTGGCATGTGGTGGGCATC 1403
 251 ValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyValTyrThrLysValSer 270
 1404 GTTAGCTGGGCTATGGCTGCGGGGCCCGAGACCCCGAGGATATACACCAAGGTCTCA 1463
 271 AlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu 284
 1464 GCCTATCTCAACTGGATCTACAATGTCTGAAGGCTGAGCTG 1505

RESULT 4

Sequence 274, Application US/10015387A

Publication No. US20030135034A1

GENERAL INFORMATION:

APPLICANT: Baker, Kevin P.

APPLICANT: Botstein, David

APPLICANT: Desnoyers, Luc

APPLICANT: Eaton, Dan L.

APPLICANT: Ferrara, Napoleone

APPLICANT: Fong, Sherman

APPLICANT: Gao, Wei-Qiang

APPLICANT: Goddard, Audrey

APPLICANT: Godowski, Paul J.

APPLICANT: Grimaldi, Christopher J.

APPLICANT: Gurney, Austin L.

APPLICANT: Hillan, Kenneth J.

APPLICANT: Pan, James

APPLICANT: Paoni, Nicholas F.

TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic

TITLE OF INVENTION: Acids Encoding the Same

FILE REFERENCE: P2830PIC54

CURRENT APPLICATION NUMBER: US/10/015,387A

Prior Application Filing Date: 2001-12-12

NUMBER OF SEQ ID NOS: 477

SEQ ID NO 274

LENGTH: 2063

TYPE: DNA

ORGANISM: Homo sapiens

-10-015-387A-274

ignment Scores:

ed. No.:

ore:

cent Similarity:

st Local Similarity:

ery Match:

:

-09-607-745-9 (1-292) x US-10-015-387A-274 (1-2063)

51 LysileValGlyGlyTyrAlaLeuAspValAspSerTrpProTrpGlnValSerIleGln 70

804 CGTGTGGTGGTGGGAGAGGCGCTCTGTGGATCTTGGCCTTGGCAGGTGAGATCCAG 863

71 TyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThrAla 90

864 TACGACAAACAGCAGCTGTGTGGAGGAGCATCTTGGACCCCACTGGTCTTACGGCA 923

91 AlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySerAsp 110

924 GCCCACTGCTTCAGGAAACATACCGATGTGTCACTGGAGGTGCGGGCAGGCTCAGAC 983

QY 111 LysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleIleGluPheAsnPro 130
 DB 984 AAACCTGGGAGCTTCCCATCCCTGGCTGTGGCAAGATCATCATCAATTCAACCCC 1043
 QY 131 MetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPheSer 150
 DB 1044 AIGTACCCCAAGACAATGACATCGCCCTCAAGAGCTGCACTCCACTCTTCTCA 1103
 QY 151 GlyThrValArgProIleCysLeuProPhePheAspGluGluLeuThrProAlaThrPro 170
 DB 1104 GGCACAGTCAGSCCCCATCTGTCTGCCCTCTTTGATGAGAGCTCACTCCAGCCACCCCA 1163
 QY 171 LeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyGlyLysMetSerAspIleLeu 190
 DB 1164 CTCTGGATCAATTGGATGGGCTTTACGAGCAGAATGGAGGAAGATGTCTGACATCTG 1223
 QY 191 LeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAlaTyrGln 210
 DB 1224 CTGCAAGGCTCAGTCCAGTCAATGACAGCACACGCTGCAATGCAGAGCATCGTACCAG 1283
 QY 211 GlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThrCys 230
 DB 1284 GGGGAAGTCACCGACAAGATGATGTGTGAGGCATCCCGAAGGGGTGTGACACCTGC 1343
 QY 231 GlnGlyAspSerGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGlyIle 250
 DB 1344 CAGGCTGACAGTGGTGGGCCCCCTGTATGATGATGATGATGATGATGATGATGATGATG 1403
 QY 251 ValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyValTyrThrLysValSer 270
 DB 1404 GTTAGCTGGGCTATGGCTGCGGGGCCCGAGACCCCGAGGATATACACCAAGGTCTCA 1463
 QY 271 AlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu 284
 DB 1464 GCCTATCTCAACTGGATCTACAATGTCTGAAGGCTGAGCTG 1505

RESULT 5

US-10-063-735-111

Sequence 111, Application US/10063735

Publication No. US2003013882A1

GENERAL INFORMATION:

APPLICANT: Eaton, Dan L.

APPLICANT: Filvaroff, Ellen

APPLICANT: Gerritsen, Mary E.

APPLICANT: Goddard, Audrey

APPLICANT: Godowski, Paul J.

APPLICANT: Grimaldi, Christopher J.

APPLICANT: Gurney, Austin L.

APPLICANT: Watanabe, Colin K.

APPLICANT: Wood, William I.

TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC

TITLE OF INVENTION: ACIDS ENCODING THE SAME

FILE REFERENCE: P3230R1C1

CURRENT APPLICATION NUMBER: US/10/063,735

CURRENT FILING DATE: 2002-05-08

Prior Application removed - See Palm or File Wrapper

NUMBER OF SEQ ID NOS: 170

SEQ ID NO 111

LENGTH: 2063

TYPE: DNA

ORGANISM: Homo Sapien

US-10-063-735-111

Alignment Scores:

Pred. No.:

Score:

Percent Similarity:

Best Local Similarity:

Query Match:

DB:

Length:

Matches:

Conservative:

Mismatches:

Indels:

Gaps:

US-09-607-745-9 (1-292) x US-10-063-735-111 (1-2063)

```
51  LysIleValGlyGlyTyrAlaLeuAspValAspSerTrpProTrpGlnValSerIleGln 70
    : : : : :
804  CGTGTGGTGGTGGGAGGAGGCTCTGTGGATTCTTGGCTTGGCAGTCCAGTCCAG 863

71  TyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThrAla 90
    : : : : :
864  TACGACAAACAGCAGCTCTGTGGAGGAGATCTTGACCCCACTGGGTCCTCAGCGCA 923

91  AlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySerAsp 110
    : : : : :
924  GCCCACTGCTTACGAAACATACCGATGTTCACTGGAGGTCGGGAGGCTCAGAC 983

111  LysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleLeuPheAsnPro 130
    : : : : :
984  AACTGGGAGCTTCCCATCCCTGGCTTGGCCCAAGATCATCATATTGAATCAACCC 1043

131  MetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPheSer 150
    : : : : :
1044  ATGTACCCCAAGACATACGATGTTCACTGGAGGTCGGGAGGCTCAGAC 983

151  GlyThrValArgProLysCysLeuProPhePheAspGluLeuThrProAlaThrPro 170
    : : : : :
1104  GGCACAGTCAGGCCCATCTGTCTGCTTCTTTGATGAGGATCACTCCAGCCACCCA 1163

171  LeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyGlyLysMetSerAspIleLeu 190
    : : : : :
1164  CTCTGGATCATTTGGATGGGCTTTACGAAGCAGATCCCGAGGGGTGTGGACACTG 1223

191  LeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAlaTrpGln 210
    : : : : :
1224  CTGAGGGCTCAGTCCAGTCTATTGACAGCACGGTGCAATGACAGTGGCATGTGG 1283

211  GlyGluValThrGluLysMetCysAlaGlyIleProGluGlyGlyValAspThrCys 230
    : : : : :
1284  GGGAGAGTCACCGAGAGATGATGTGTGAGGATCCCGAGGGGTGTGGACACTG 1343

231  GlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGlyIle 250
    : : : : :
1344  CAGGTGACAGTGGTGGGCCCCGTGATGATACCAATCTGACAGTGGCATGTGGG 1403

251  ValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyValTyrThrLysValSer 270
    : : : : :
1404  GTTAGCTGGGCTATGGCTCGGGGCCCCGAGCACCCGAGGATATACACCAAGGTCTCA 1463

271  AlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu 284
    : : : : :
1464  GCCTATCTCACTGGATCTACATGTTCTGGAAGGCTGAGCTG 1505
```

RESULT 6

```
US-10-006-130A-274
Sequence 274, Application US/10006130A
Publication No. US20030148375A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Botstein, David
APPLICANT: Desnoyers, Luc
APPLICANT: Eaton, Dan L.
APPLICANT: Ferrara, Napoleone
APPLICANT: Fong, Sherman
APPLICANT: Gao, Wei-Qiang
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Grimaldi, Christopher J.
APPLICANT: Gurney, Austin L.
APPLICANT: Hillan, Kenneth J.
APPLICANT: Pan, James
APPLICANT: Paoni, Nicholas F.
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
TITLE OF INVENTION: Acids Encoding the Same
FILE REFERENCE: P2830PIC7
CURRENT APPLICATION NUMBER: US/10/006,130A
CURRENT FILING DATE: 2002-03-19
Prior Application removed - See File Wrapper or Palm
```

```
; NUMBER OF SEQ ID NOS: 477
; SEQ ID NO 274
; LENGTH: 2063
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-006-130A-274
```

Alignment Scores:

```
Pred. No.: 1-6e-152 Length: 2063
Score: 1266.00 Matches: 228
Percent Similarity: 98.28% Conservatives: 2
Best Local Similarity: 97.44% Mismatches: 4
Query Match: 79.12% Indels: 0
DB: 12 Gaps: 0
```

US-09-607-745-9 (1-292) x US-10-006-130A-274 (1-2063)

```
QY 51 LysIleValGlyGlyTyrAlaLeuAspValAspSerTrpProTrpGlnValSerIleGln 70
    : : : : :
Db 804 CGTGTGGTGGTGGGAGGAGGCTCTGTGGATTCTTGGCTTGGCAGTCCAGTCCAG 863

QY 71 TyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThrAla 90
    : : : : :
Db 864 TACGACAAACAGCAGCTCTGTGGAGGAGATCTTGACCCCACTGGGTCCTCAGCGCA 923

QY 91 AlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySerAsp 110
    : : : : :
Db 924 GCCCACTGCTTACGAAACATACCGATGTTCACTGGAGGTCGGGAGGCTCAGAC 983

QY 111 LysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleGluPheAsnPro 130
    : : : : :
Db 984 AACTGGGAGCTTCCCATCCCTGGCTTGGCCCAAGATCATCATATTGAATCAACCC 1043

QY 131 MetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPheSer 150
    : : : : :
Db 1044 ATGTACCCCAAGACATACGATGTTCACTGGAGGTCGGGAGGCTCAGAC 983

QY 151 GlyThrValArgProLysCysLeuProPhePheAspGluLeuThrProAlaThrPro 170
    : : : : :
Db 1104 GGCACAGTCAGGCCCATCTGTCTGCTTCTTTGATGAGGATCACTCCAGCCACCCA 1163

QY 171 LeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyGlyLysMetSerAspIleLeu 190
    : : : : :
Db 1164 CTCTGGATCATTTGGATGGGCTTTACGAAGCAGATCCCGAGGGGTGTGGACACTG 1223

QY 191 LeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAlaTrpGln 210
    : : : : :
Db 1224 CTGAGGGCTCAGTCCAGTCTATTGACAGCACGGTGCAATGACAGTGGCATGTGG 1283

QY 211 GlyGluValThrGluLysMetCysAlaGlyIleProGluGlyGlyValAspThrCys 230
    : : : : :
Db 1284 GGGAGAGTCACCGAGAGATGATGTGTGAGGATCCCGAGGGGTGTGGACACTG 1343

QY 231 GlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGlyIle 250
    : : : : :
Db 1344 CAGGTGACAGTGGTGGGCCCCGTGATGATACCAATCTGACAGTGGCATGTGGG 1403

QY 251 ValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyValTyrThrLysValSer 270
    : : : : :
Db 1404 GTTAGCTGGGCTATGGCTCGGGGCCCCGAGCACCCGAGGATATACACCAAGGTCTCA 1463

QY 271 AlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu 284
    : : : : :
Db 1464 GCCTATCTCACTGGATCTACATGTTCTGGAAGGCTGAGCTG 1505
```

RESULT 7

```
US-10-199-672-329
; Sequence 329, Application US/10199672
; Publication No. US20030148442A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
```

APPLICANT: Goddard,Audrey
 APPLICANT: Godowski,Paul J.
 APPLICANT: Gurney,Austin L.
 APPLICANT: Pan,James
 APPLICANT: Smith,Victoria
 APPLICANT: Watanabe,Colin K.
 APPLICANT: Wood,William I.
 APPLICANT: Zhang,Zemin

TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC ACIDS ENCODING THE SAME

CURRENT APPLICATION NUMBER: US/10/199,672
 CURRENT FILING DATE: 2002-07-18
 PRIOR APPLICATION NUMBER: US/10/052,586
 PRIOR FILING DATE: 2002-01-15
 PRIOR APPLICATION NUMBER: 60/059263
 PRIOR FILING DATE: 1997-09-18
 PRIOR APPLICATION NUMBER: 60/059266
 PRIOR FILING DATE: 1997-09-18
 PRIOR APPLICATION NUMBER: 60/062250
 PRIOR FILING DATE: 1997-10-17
 PRIOR APPLICATION NUMBER: 60/063120
 PRIOR FILING DATE: 1997-10-24
 PRIOR APPLICATION NUMBER: 60/063121
 PRIOR FILING DATE: 1997-10-24
 PRIOR APPLICATION NUMBER: 60/063486
 PRIOR FILING DATE: 1997-10-21
 PRIOR APPLICATION NUMBER: 60/063540
 PRIOR FILING DATE: 1997-10-28
 PRIOR APPLICATION NUMBER: 60/063541
 PRIOR FILING DATE: 1997-10-28
 PRIOR APPLICATION NUMBER: 60/063544
 PRIOR FILING DATE: 1997-10-28
 Remaining Prior Application data removed - See File Wrapper or PALM.

NUMBER OF SEQ ID NOS: 612
 SEQ ID NO 329

LENGTH: 2063
 TYPE: DNA
 ORGANISM: Homo Sapien
 -10-199-672-329

Alignment Scores:	Seq No.:	Length:	Matches:	Mismatches:	Gaps:
1.6e-152		2063			
1266.00		228			
98.29%		Conservative:	2		
97.44%		Mismatches:	4		
79.12%		Indels:	0		
12		Gaps:	0		

-09-607-745-9 (1-292) x US-10-199-672-329 (1-2063)

51 LysileValGlyGlyTyraAlaLeuaspValaspSerTrpProTrpGlnValserIleGln 70
 ::::
 804 CGTGTGGTGTTGGAGGAGGCCTCTGTGATTCCTGGCCCTGGCAGGTACAGATCCAG 863

71 TyrAspLysGlnHisValCysGlyGlySerlleLeuaspProHisTrpValLeuThrAla 90
 ::::
 864 TAGCACAAACAGCACGCTGTGTGGAGGAGCATCTCGACCACCCCACTGGTCTCTCACGCCA 923

91 AlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySerAsp 110
 ::::
 924 GCCACATGCTTCAGAAACATACCGATGTGTTCACCTGAAGAGGTGCGGGCAGGTCTGAC 983

111 LysLeuGlySerPheProSerLeuAlaValAlaLysillellelleGluPheAsnPro 130
 ::::
 984 AAACCTGGCAGCTTCCCATTCTGGCTGTGGCCAGATCATCATTCATTGAATTCACCCC 1043

131 MetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPheSer 150
 ::::
 1044 ATGTACCCCCAAAGACAATGACATCGCCCTCATGAAGCTGAGTTCACACTTCTCA 1103

151 GlyThrValArgProIleCysLeuProPhePheaspGluLeuLeuThrProAlaThrPro 170
 ::::
 ::::

1	1	PRIOR APPLICATION NUMBER: 60/101741
2	2	PRIOR FILING DATE: 1998-09-24
3	3	PRIOR APPLICATION NUMBER: 60/101743
4	4	PRIOR FILING DATE: 1998-09-24
5	5	PRIOR APPLICATION NUMBER: 60/101915
6	6	PRIOR FILING DATE: 1998-09-24
7	7	PRIOR APPLICATION NUMBER: 60/101916
8	8	PRIOR FILING DATE: 1998-09-24
9	9	PRIOR APPLICATION NUMBER: 60/102207
10	10	PRIOR FILING DATE: 1998-09-29
11	11	PRIOR APPLICATION NUMBER: 60/102240
12	12	PRIOR FILING DATE: 1998-09-29
13	13	PRIOR APPLICATION NUMBER: 60/102307
14	14	PRIOR FILING DATE: 1998-09-29
15	15	PRIOR APPLICATION NUMBER: 60/102330
16	16	PRIOR FILING DATE: 1998-09-29
17	17	PRIOR APPLICATION NUMBER: 60/102331
18	18	PRIOR FILING DATE: 1998-09-29
19	19	PRIOR APPLICATION NUMBER: 60/102484
20	20	PRIOR FILING DATE: 1998-09-30
21	21	PRIOR APPLICATION NUMBER: 60/102487
22	22	PRIOR FILING DATE: 1998-09-30
23	23	PRIOR APPLICATION NUMBER: 60/102570
24	24	PRIOR FILING DATE: 1998-09-30
25	25	PRIOR APPLICATION NUMBER: 60/102571
26	26	PRIOR FILING DATE: 1998-09-30
27	27	PRIOR APPLICATION NUMBER: 60/102684
28	28	PRIOR FILING DATE: 1998-10-01
29	29	PRIOR APPLICATION NUMBER: 60/102687
30	30	PRIOR FILING DATE: 1998-10-01
31	31	PRIOR APPLICATION NUMBER: 60/102965
32	32	PRIOR FILING DATE: 1998-10-02
33	33	PRIOR APPLICATION NUMBER: 60/103258
34	34	PRIOR FILING DATE: 1998-10-06
35	35	PRIOR APPLICATION NUMBER: 60/103314
36	36	PRIOR FILING DATE: 1998-10-07
37	37	PRIOR APPLICATION NUMBER: 60/103315
38	38	PRIOR FILING DATE: 1998-10-07
39	39	PRIOR APPLICATION NUMBER: 60/103328
40	40	PRIOR FILING DATE: 1998-10-07
41	41	PRIOR APPLICATION NUMBER: 60/103395
42	42	PRIOR FILING DATE: 1998-10-07
43	43	PRIOR APPLICATION NUMBER: 60/103396
44	44	PRIOR FILING DATE: 1998-10-07
45	45	PRIOR APPLICATION NUMBER: 60/103401
46	46	PRIOR FILING DATE: 1998-10-07
47	47	PRIOR APPLICATION NUMBER: 60/103449
48	48	PRIOR FILING DATE: 1998-10-06
49	49	PRIOR APPLICATION NUMBER: 60/103633
50	50	PRIOR FILING DATE: 1998-10-08
51	51	PRIOR APPLICATION NUMBER: 60/103678
52	52	PRIOR FILING DATE: 1998-10-08
53	53	PRIOR APPLICATION NUMBER: 60/103679
54	54	PRIOR FILING DATE: 1998-10-08
55	55	PRIOR APPLICATION NUMBER: 60/103711
56	56	PRIOR FILING DATE: 1998-10-08
57	57	PRIOR APPLICATION NUMBER: 60/104257
58	58	PRIOR FILING DATE: 1998-10-14
59	59	PRIOR APPLICATION NUMBER: 60/104987
60	60	PRIOR FILING DATE: 1998-10-20
61	61	PRIOR APPLICATION NUMBER: 60/105000
62	62	PRIOR FILING DATE: 1998-10-20
63	63	PRIOR APPLICATION NUMBER: 60/105002
64	64	PRIOR FILING DATE: 1998-10-20
65	65	PRIOR APPLICATION NUMBER: 60/105104
66	66	PRIOR FILING DATE: 1998-10-21
67	67	PRIOR APPLICATION NUMBER: 60/105169
68	68	PRIOR FILING DATE: 1998-10-22
69	69	PRIOR APPLICATION NUMBER: 60/105266
70	70	PRIOR FILING DATE: 1998-10-22
71	71	PRIOR APPLICATION NUMBER: 60/105693
72	72	PRIOR FILING DATE: 1998-10-26
73	73	PRIOR APPLICATION NUMBER: 60/105694

PRIOR FILING DATE: 1998-10-26
PRIOR APPLICATION NUMBER: 60/105807
PRIOR FILING DATE: 1998-10-27
PRIOR APPLICATION NUMBER: 60/105881
PRIOR FILING DATE: 1998-10-27
PRIOR APPLICATION NUMBER: 60/105882
PRIOR FILING DATE: 1998-10-27
PRIOR APPLICATION NUMBER: 60/106023
PRIOR FILING DATE: 1998-10-28

Alignment Scores:
Pred. No.: 1,6e-152 Length: 2063
Score: 1266.00 Matches: 228
Percent Similarity: 98.29% Conservative: 2
Best Local Similarity: 97.44% Mismatches: 4
Query Match: 79.12% Indels: 0
Gaps: 12

US-09-607-745-9 (1-292) x US-10-006-172A-274 (1-2063)

51 LysIleValGlyGlyTyrAlaLeuAspValAspSerTrpProTrpGlnValSerIleGln 70
804 CCGTGGTGGGGAGAGAGCCCTCTGGGATCTTGGCTTGGCAGTCCAGTCCAG 863
71 TyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThrAla 90
864 TACGACAAACACACAGCTGTGGAGGAGAGATCTGGACCCCACTGGGTCTCCAGGCA 923
91 AlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySerAsp 110
924 GCCCAGTCTTCAGGAACATACCGATGTGTTCACTGGAAGTGGGGCAGGCCTCAGAC 983
111 LysLeuGlySerPheProSerLeuAlaValAlaLysIleIlellelleGluPheAsnPro 130
984 AAACCTGGGAGCTTCCATCCCTGGCTGGGCGAGATCATCATTAATTAACCC 1043
131 MetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPheSer 150
1044 ATGTACCCCAAGACAAAGACATGCGCCCTCATGAAGCTGAGTCCCACTCACTTTCTCA 1103
151 GlyThrValArgProIleCysLeuProPhePheAspGluGluLeuThrProAlaThrPro 170
1104 GGCACAGTCAGGCCCATCTGTGCGCTTCTTGTAGAGAGCTCACTCCAGCCACCCA 1163
171 LeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyGlyLysMetSerAspIleLeu 190
1164 CTCTGGATCATTTGGATGGGCTTTACAGACGAGATGGAGGAGATGCTCAGCATACTG 1223
191 LeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAlaTyrGln 210
1224 CTGACGGGCTGAGTCCAGTCTATTGACAGCACACGGTGCATGACAGATCGTACCCAG 1283
211 GlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThrCys 230
1284 GGGGAAGTCACCGAAGATGATGTGTCAGGCATCCCGGAAGGGGGTGGACACCTGC 1343
231 GlnGlyAspSerGlyProLeuMetTyrGlnSerAspGlnTrpHisValGlyIle 250
1344 CAGGCTGACAGTGGGGGGCCCTGATGATACCAATCTGACCACTGAGTGGGCGATC 1403
251 ValSerTrpGlyTyrGlyCysGlyProSerThrProGlyValTyrThrLysValSer 270
1404 GTTACTGGGGTATGGCTGGGGGGCCCGACACCCCAAGAGATATACCAAGGTCTCA 1463
271 AlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu 284
1464 GCCTATCTCAACTGATCTACAAATGCTTGGAGGCTGAGCTG 1505

RESULT 9

-10-187-749-329

Sequence 329, Application US/10187749

Publication No. US2003015036A1

GENERAL INFORMATION:

APPLICANT: Baker, Kevin P.
APPLICANT: Chen, Jian
APPLICANT: Desnoyers, Luc
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Pan, James
APPLICANT: Smith, Victoria
APPLICANT: Watanabe, Colin K.
APPLICANT: Wood, William I.
APPLICANT: Zhang, Zemin

TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
ACIDS ENCODING THE SAME

FILE REFERENCE: P3430RIC1

CURRENT APPLICATION NUMBER: US/10/187,749

PRIOR FILING DATE: 2002-07-01

PRIOR APPLICATION NUMBER: US/10/052,586

PRIOR FILING DATE: 2002-01-15

PRIOR APPLICATION NUMBER: 60/059263

PRIOR FILING DATE: 1997-09-18

PRIOR APPLICATION NUMBER: 60/059266

PRIOR FILING DATE: 1997-09-18

PRIOR APPLICATION NUMBER: 60/062250

PRIOR FILING DATE: 1997-10-17

PRIOR APPLICATION NUMBER: 60/063120

PRIOR FILING DATE: 1997-10-24

PRIOR APPLICATION NUMBER: 60/063121

PRIOR FILING DATE: 1997-10-24

PRIOR APPLICATION NUMBER: 60/063486

PRIOR FILING DATE: 1997-10-21

PRIOR APPLICATION NUMBER: 60/063540

PRIOR FILING DATE: 1997-10-28

PRIOR APPLICATION NUMBER: 60/063541

PRIOR FILING DATE: 1997-10-28

PRIOR APPLICATION NUMBER: 60/063544

PRIOR FILING DATE: 1997-10-28

Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 612
SEQ ID NO 329

LENGTH: 2063

TYPE: DNA

ORGANISM: Homo Sapien

US-10-187-749-329

Alignment Scores:

Pred. No.: 1,6e-152 Length: 2063
Score: 1266.00 Matches: 228
Percent Similarity: 98.29% Conservative: 2
Best Local Similarity: 97.44% Mismatches: 4
Query Match: 79.12% Indels: 0
Gaps: 12

US-09-607-745-9 (1-292) x US-10-187-749-329 (1-2063)

QY 51 LysIleValGlyGlyTyrAlaLeuAspValAspSerTrpProTrpGlnValSerIleGln 70

Db 804 CCGTGGTGGGGAGAGAGCCCTCTGGGATCTTGGCTTGGCAGTCCAGTCCAG 863

QY 71 TyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThrAla 90

Db 864 TACGACAAACACACAGCTGTGGAGGAGAGATCTTGGACCCCACTGGGTCTCCAGGCA 923

QY 91 AlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySerAsp 110

Db 924 GCCCAGTCTTCAGGAACATACCGATGTGTTCACTGGAAGTGGGGCAGGCTCAGAC 983

QY 111 LysLeuGlySerPheProSerLeuAlaValAlaLysIleIlellelleGluPheAsnPro 130

Db 984 AAACCTGGGAGCTTCCATCCCTGGCTGGGCGAGATCATCATTAATTAACCC 1043

QY 131 MetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPheSer 150

Db 1044 ATGTACCCCAAGACAAAGACATGCGCCCTCATGAAGTGGAGTCCCACTCACTTTCTCA 1103

151 GlyThrValArgProIleCysLeuProPhePheAspGluLeuThrProAlaThrPro 170
1104 GGCACAGTCCAGGCCCATCTGCTGCGCTTCTTTGATCAGGAGCTCACTCCAGCCACCCCA 1163
171 LeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyLysMetSerAspIleLeu 190
1164 CTCCTGATCATTTGGATGGGCTTTACGAGCAGAATGGAGGGAAGATGCTGCATCACTG 1223
191 LeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAlaThrGln 210
1224 CTCGAGGCGTCACTCCAGTCACTGACAGACACGCTGCTACACACAGATGGCTACCG 1283
211 GlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThrCys 230
1284 GGGGAAGTCCAGGAGAGATGATGTGTGAGGATCCCGGAGGGGCTGTGACACCTGC 1343
231 GlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValGlyIle 250
1344 CAGGTGACAGTGGTGGGCCCCCTGATGTACCAATCTGACACAGTGGCATGTGGTGGCATC 1403
251 ValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyValTyrThrLysValSer 270
1404 GTAGCTGGGCTATGGCTGCGGGGCGCCGAGCACCCCGAGGATACACCAAGTCTCA 1463
271 AlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu 284
1464 GCCTATCTCAACTGGATCTACAATGTCTGGAAGGCTGAGCTG 1505

RESULT 10

US-10-194-457-329
Sequence 329, Application US/10194457
Publication No. US20030153037A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Chen, Jian
APPLICANT: Desnoyers, Luc
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Pan, James
APPLICANT: Smith, Victoria
APPLICANT: Watanabe, Colin K.
APPLICANT: Wood, William I.
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
TITLE OF INVENTION: ACIDS ENCODING THE SAME
FILE REFERENCE: P3430R1C296
CURRENT APPLICATION NUMBER: US/10/194,457
CURRENT FILING DATE: 2002-07-11
PRIOR APPLICATION NUMBER: 10/052586
PRIOR FILING DATE: 2002-01-15
PRIOR APPLICATION NUMBER: 60/059263
PRIOR FILING DATE: 1997-09-18
PRIOR APPLICATION NUMBER: 60/059266
PRIOR FILING DATE: 1997-09-18
PRIOR APPLICATION NUMBER: 60/062250
PRIOR FILING DATE: 1997-10-17
PRIOR APPLICATION NUMBER: 60/063120
PRIOR FILING DATE: 1997-10-24
PRIOR APPLICATION NUMBER: 60/063121
PRIOR FILING DATE: 1997-10-24
PRIOR APPLICATION NUMBER: 60/063486
PRIOR FILING DATE: 1997-10-21
PRIOR APPLICATION NUMBER: 60/063540
PRIOR FILING DATE: 1997-10-28
PRIOR APPLICATION NUMBER: 60/063541
PRIOR FILING DATE: 1997-10-28
PRIOR APPLICATION NUMBER: 60/063544
PRIOR FILING DATE: 1997-10-28
Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 612

SEQ ID NO 329

LENGTH: 2063
TYPE: DNA
ORGANISM: Homo Sapien
US-10-194-457-329
Alignment Scores:
Pred. No.: 1.6e-152 Length: 2063
Score: 1266.00 Matches: 228
Percent Similarity: 98.29% Conservative: 2
Best Local Similarity: 97.44% Mismatches: 4
Query Match: 79.12% Indels: 0
DB: 12 Gaps: 0

US-09-607-745-9 (1-292) x US-10-194-457-329 (1-2063)

QY 51 LysIleValGlyGlyTyrAlaLeuAspValAspSerTrpProTrpGlnValSerIleGln 70
Db 804 CCGTGGTGGTGGGAGGAGGCGCTCTGTGGATTCTTGGCCCTTGGCAGGTCCAGTCCAG 863
QY 71 TyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThrAla 90
Db 864 TACGACAAACAGCAGCTCTGTGGAGGAGCATCTTGACCCCTTGGTCTCTCAGGCA 923
QY 91 AlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySerAsp 110
Db 924 GCCCACTGCTTCAGGAACATACCGATGTGTCACTGGAAGGTGGGCGAGGCTCAGAC 983
QY 111 LysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleGluPheAsnPro 130
Db 984 AAACCTGGGAGCTTCCCATCCCTGGCTGTGCGCAAGATCATCATTAATGAATCAACCCC 1043
QY 131 MetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPheSer 150
Db 1044 ATGTACCCCAAGACAATGACATCGCCCTCATGAAGTGCAGTTCCTCCTTCTTCTCA 1103
QY 151 GlyThrValArgProIleCysLeuProPhePheAspGluLeuThrProAlaThrPro 170
Db 1104 GGCACAGTCAGGCCCATCTGCTGCCCTTCTTTGATGAGGAGTCACTCCAGCCACCCCA 1163
QY 171 LeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyLysMetSerAspIleLeu 190
Db 1164 CTCCTGATCATTTGGATGGGCTTTACGAAGCAGAGATGAGGGAAGATGCTGCATACTG 1223
QY 191 LeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAlaThrGln 210
Db 1224 CTCGAGGCGTCACTCCAGTCACTGACAGACACAGTGCATGACAGATGCGTACCCAG 1283
QY 211 GlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThrCys 230
Db 1284 GGGGAAGTCCAGGAGAGATGATGTGTGAGGATCCCGGAGGGGCTGTGACACCTGC 1343
QY 231 GlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGlyIle 250
Db 1344 CAGGTGACAGTGGTGGGCCCCCTGATGTACCAATCTGACCAAGTGGCATGTGGTGGCATC 1403
QY 251 ValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyValTyrThrLysValSer 270
Db 1404 GTTAGCTGGGCTATGGCTGCGGGGCGCCGAGCACCCCGAGGATATACACCAAGTCTCA 1463
QY 271 AlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu 284
Db 1464 GCCTATCTCAACTGGATCTACAATGTCTGGAAGGCTGAGCTG 1505

RESULT 11

US-10-184-642-329
Sequence 329, Application US/10184642
Publication No. US20030157635A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Chen, Jian
APPLICANT: Desnoyers, Luc
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.

APPLICANT: Gurney,Austin L.
 APPLICANT: Pan,James
 APPLICANT: Smith,Victoria
 APPLICANT: Watanabe,Colin K.
 APPLICANT: Wood,William I.
 APPLICANT: Zhang,Zemin

TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
 FILE REFERENCE: P3430R1C194

CURRENT APPLICATION NUMBER: US/10/184,642

Prior Application removed - See File Wrapper or Palm

SEQUENCE ID NOS: 612

LENGTH: 2063

TYPE: DNA

ORGANISM: Homo Sapien

US-10-184-642-329

Alignment Scores:

Pred. No.:	1-6e-152	Length:	2063
Score:	1266.00	Matches:	228
Percent Similarity:	98.29%	Conservative:	2
Best Local Similarity:	97.44%	Mismatches:	4
Query Match:	79.12%	Indels:	0
Gaps:	12		

US-09-607-745-9 (1-292) x US-10-184-642-329 (1-2063)

51	LysIleValGlyGlyTyrAlaLeuAspValAspSerTrpProTropGlnValSerIleGln	70
804	CTGTGGTGGTGGGAGGAGGCTCTGTGGATCTTGGCTTGGCAGTCAGATCCAG	863
71	TyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThrAla	90
864	TACGACAAACAGACACGCTCTGTGGAGGAGCATCTGGACCCCATCTGCTCCACGGCA	923
91	AlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySerAsp	110
924	GCCACTGCTTCAGGAACATACCGGATGTTCACTGGAAGGTGGGGCAGCTCAGAC	983
111	LysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleIleGluPheAsnPro	130
984	AAACTGGGCGAGCTTCCATCTCTGGCTGGGCGGAGCATCATCATCTTGAATTCACCCC	1043
131	MetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPheSer	150
1044	ATGTACCCCAAGACAAATGACATCGCCCTCATGAAGCTGCAGTCCACTCTCTCTCA	1103
151	GlyThrValArgProIleCysLeuProPhePheAspGluGluLeuThrProAlaThrPro	170
1104	GGCAGAGTCAGGCCCATCTGTCTGCCCTCTTTGATGAGGAGCTCACTCCAGCCACCCA	1163
171	LeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyLysMetSerAspIleLeu	190
1164	CTCTGGATCATTTGGATGGGCTTACGAGCAGATGAGGAGGAGATGCTGACATCTG	1223
191	LeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAlaTyrGln	210
1224	CTGCAGGGGTGATCCAGTCAATGACAGCACACGCTGCAATGCAGACGATCGTACCA	1283
211	GlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThrCys	230
1284	GGGGAAGTCACGAGAGATGATGTGTGAGGAGATCCCGGAGGGGGGTGTGACACCTGC	1343
231	GlnGlyAspSerGlyLysProLeuMetTyrGlnSerAspGlnTrpHisValValGlyIle	250
1344	CAGGGTGACAGTGGTGGGCTCTGATGTACCAATCTGACCATGGCATGTGGTGGGCATC	1403
251	ValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyValTyrThrLysValSer	270
1404	GTAGCTGGGCTATGTGGTGGGCGGCGGAGCACCCGAGGAGTATACACCAAGGTCTCA	1463

QY 271 AlaTyrIleuAsnTrpIleTyrIleValTrpLysAlaGluLeu 284
 Db 1464 GCCTATCTCAACTGGATCTACAATGCTCGAAGGCTGAGCTG 1505

RESULT 12

US-10-196-747-329

Sequence 329, Application US/10196747

Publication No. US20030162250A1

GENERAL INFORMATION:

APPLICANT: Baker, Kevin P.

APPLICANT: Chen, Jian

APPLICANT: Desnoyers, Luc

APPLICANT: Goddard, Audrey

APPLICANT: Godowski, Paul J.

APPLICANT: Gurney, Austin L.

APPLICANT: Pan, James

APPLICANT: Smith, Victoria

APPLICANT: Watanabe, Colin K.

APPLICANT: Wood, William I.

APPLICANT: Zhang, Zemin

TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC

FILE REFERENCE: P3430R1C346

CURRENT APPLICATION NUMBER: US/10/196,747

CURRENT FILING DATE: 2002-07-16

Prior Application removed - See File Wrapper or Palm

NUMBER OF SEQ ID NOS: 612

SEQ ID NO 329

LENGTH: 2063

TYPE: DNA

ORGANISM: Homo Sapien

US-10-196-747-329

Alignment Scores:

Pred. No.:	1-6e-152	Length:	2063
Score:	1266.00	Matches:	228
Percent Similarity:	98.29%	Conservative:	2
Best Local Similarity:	97.44%	Mismatches:	4
Query Match:	79.12%	Indels:	0
Gaps:	12		

US-09-607-745-9 (1-292) x US-10-196-747-329 (1-2063)

QY	51	LysIleValGlyGlyTyrAlaLeuAspValAspSerTrpProTropGlnValSerIleGln	70
Db	804	CTGTGGTGGTGGGAGGAGGCTCTGTGGATCTTGGCTTGGCAGTCAGATCCAG	863
QY	71	TyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThrAla	90
Db	864	TACGACAAACAGACACGCTCTGTGGAGGAGCATCTGGACCCCATCTGCTCCACGGCA	923
QY	91	AlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySerAsp	110
Db	924	GCCACTGCTTCAGGAACATACCGATGTTCACTGGAAGGTGGGGCAGGCTCAGAC	983
QY	111	LysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleIleGluPheAsnPro	130
Db	984	AAACTGGGCGAGCTTCCATCTCTGGCTGGGCGGAGCATCTGGACCCCATCTGCTCCAC	1043
QY	131	MetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPheSer	150
Db	1044	ATGTACCCCAAGACAAATGACATCGCCCTCATGAAGCTGCAGTCCACTCTCTCTCA	1103
QY	151	GlyThrValArgProIleCysLeuProPhePheAspGluGluLeuThrProAlaThrPro	170
Db	1104	GGCAGAGTCAGGCCCATCTGTCTGCCCTCTTTGATGAGGAGCTCACTCCAGCCACCCA	1163
QY	171	LeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyLysMetSerAspIleLeu	190
Db	1164	CTCTGGATCATTTGGATGGGCTTACGAGCAGATGAGGAGGAGATGCTGACATCTG	1223
QY	191	LeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAlaTyrGln	210

b 1224 CTGACGGCTCAGTCCAGGTCATTGACAGACACGGTGCATGACAGCATGCTACACAG 1283
y 211 GlyGluValThrGluLysMetMetCysAlaGlyLeuProGluGlyGlyValAspThrCys 230
b 1284 GGGGAAGTCACCGAGAAGATGATGTGTGACAGCATCCCGAAGGGGTGTGGACCTGC 1343
y 231 GlnGlyAspSerGlyGlyProLeuMetTyxClnSerAspGlnThrHisValValGlyLe 250
b 1344 CAGGGTGACAGTGGTGGGCCCCCTCATGTACCAATCTGACCAAGTGGCATGTGTGGGCATC 1403
y 251 ValSerTrpGlyTyrrGlyCysGlyGlyProSerThrProGlyValTyrrLysValSer 270
b 1404 GTTAGCTGGGCTTAGCTGGGGGGCCCCGAGCACCACGAGGATATACCAAGGTCTCA 1463
y 271 AlaTyrrLeuAsnTrpIleTyrrAsnValTrpLysAlaGluLeu 284
b 1464 GCCTATCTCAACTGGATCTCAATGTCTGGAAGGCTGAGCTG 1505

RESULT 13

S-10-015-392A-274
Sequence 274, Application US/10015392A
Publication No. US20030166901A1

GENERAL INFORMATION:

APPLICANT: Baker, Kevin P.
APPLICANT: Botstein, David
APPLICANT: Desnoyers, Luc
APPLICANT: Eaton, Dan I.
APPLICANT: Ferrara, Napoleone
APPLICANT: Fong, Sherman
APPLICANT: Gao, Wei-Qiang
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Grimaldi, Christopher J.
APPLICANT: Gurney, Austin L.
APPLICANT: Hillan, Kenneth J.
APPLICANT: Pan, James
APPLICANT: Pisoni, Nicholas F.

TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic

FILE OF INVENTION: Acids Encoding the Same

FILE REFERENCE: P2830P1C58

CURRENT APPLICATION NUMBER: US/10/015,392A

CURRENT FILING DATE: 2001-12-12

PRIOR APPLICATION NUMBER: 60/098716

PRIOR FILING DATE: 1998-09-01

PRIOR APPLICATION NUMBER: 60/098723

PRIOR FILING DATE: 1998-09-01

PRIOR APPLICATION NUMBER: 60/098749

PRIOR FILING DATE: 1998-09-01

PRIOR APPLICATION NUMBER: 60/098750

PRIOR FILING DATE: 1998-09-01

PRIOR APPLICATION NUMBER: 60/098803

PRIOR FILING DATE: 1998-09-02

PRIOR APPLICATION NUMBER: 60/098821

PRIOR FILING DATE: 1998-09-02

PRIOR APPLICATION NUMBER: 60/098843

PRIOR FILING DATE: 1998-09-02

PRIOR APPLICATION NUMBER: 60/099536

PRIOR FILING DATE: 1998-09-09

PRIOR APPLICATION NUMBER: 60/099596

PRIOR FILING DATE: 1998-09-09

PRIOR APPLICATION NUMBER: 60/099598

PRIOR FILING DATE: 1998-09-09

Remaining Prior Application data removed - See File Wrapper or PALM.

NUMBER OF SEQ ID NOS: 477

SEQ ID NO 274

LENGTH: 2063

TYPE: DNA

ORGANISM: Homo sapiens

S-10-015-392A-274

Alignment Scores:

red. No.: 1,6e-152

core: 1266.00

Length: 2063

Matches: 228

Percent Similarity: 98.29% Conservative: 2
Best Local Similarity: 97.44% Mismatches: 4
Query Match: 79.12% Indels: 0
DB: 12 Gaps: 0

US-09-607-745-9 (1-292) x US-10-015-392A-274 (1-2063)

QY 51 LysIleValGlyGlyTyrrAlaLeuAspValAspSerTrpProTrpGlnValSerIleGln 70
Db 804 CGTGTGTGGTGGGAGGAGGCGCTCTGTGGATCTTGGCTTGGCAGGTGACATCCAG 863
QY 71 TyrAspIleGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThrAla 90
Db 864 TACGACAAACAGCAGCATCTGTGTGAGGAGCATCTCTGACCCCCCTGGTCCCTCACGGCA 923
QY 91 AlaHisCysPheArgGlyHisThrAspValPheAsnTrpLysValArgAlaGlySerAsp 110
Db 924 GCCCACTGCTTCAGGAAACATACCGATGTGTCACTGGAAGGTGCGGCGCAGGCTCAGAC 983
QY 111 LysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleIleGluPheAsnPro 130
Db 984 AAACCTGGCAGCTTCCCATCCCTGGCTGTGGCCAGATCATCATTAATTCAACCCC 1043
QY 131 MetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPheSer 150
Db 1044 ATGTACCCCAAGACAAATGACATCGCCCTCATGAGCTGAGTCCCACTCACTTTCTCA 1103
QY 151 GlyThrValArgProIleCysLeuProPheAspGluGluLeuThrProAlaThrPro 170
Db 1104 GGCACAGTCAGGCCCACTCTGTCTGCCCTTCTTTGATGAGGAGCTCACTCCAGCCACCCCA 1163
QY 171 LeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyGlyLysMetSerAspIleLeu 190
Db 1164 CTCTGATCATTTGGATGGGCTTTACGAGACAGATGGAGGAGATGCTGCACACTCTG 1223
QY 191 LeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAlaTyrrGln 210
Db 1224 CTGACGGCTGATGTCAGGTTCATTGACAGCACACGGTGCAATGCAGACGATGCTACCAG 1283
QY 211 GlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThrCys 230
Db 1284 GGGGAAGTCACCGAGAAGATGATGTGTGACAGCATCCCGAAGGGGTGTGGACCTGC 1343
QY 231 GlnGlyAspSerGlyGlyProLeuMetTyrrGlnSerAspGlnTrpHisValValGlyLe 250
Db 1344 CAGGGTGACAGTGGTGGGCCCCCTGATGTACCAATCTGACCAAGTGGCATGTGTGGGCATC 1403
QY 251 ValSerTrpGlyTyrrGlyCysGlyGlyProSerThrProGlyValTyrrLysValSer 270
Db 1404 GTTAGCTGGGCTATGCTGCGGGGGCCCCGAGCACCAGGAGATATACCAAGGTCTCA 1463
QY 271 AlaTyrrLeuAsnTrpIleTyrrAsnValTrpLysAlaGluLeu 284
Db 1464 GCCTATCTCAACTGGATCTCAATGTCTGGAAGGCTGAGCTG 1505

RESULT 14

US-10-017-253A-274

Sequence 274, Application US/10017253A

Publication No. US20030166055A1

GENERAL INFORMATION:

APPLICANT: Baker, Kevin P.
APPLICANT: Botstein, David
APPLICANT: Desnoyers, Luc
APPLICANT: Eaton, Dan I.
APPLICANT: Ferrara, Napoleone
APPLICANT: Fong, Sherman
APPLICANT: Gao, Wei-Qiang
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Grimaldi, Christopher J.
APPLICANT: Gurney, Austin L.
APPLICANT: Hillan, Kenneth J.
APPLICANT: Pan, James

Alignment Scores:	1.6e-152	Length:	2063
ed. NO.:	1266.00	Matches:	228
ore:	98.29%	Conservative:	2
Percent Similarity:	97.44%	Mismatches:	4
st Local Similarity:	79.12%	Indels:	0
ery Match:	12	Gaps:	0
:			
(-09-607-745-9 (1-292) x US-10-017-253A-274 (1-2063))			
51	LysLeValGlyGlyTyAlaLeuAspValAspSerTrpProTrpGlnValSerIleGln	70	
804	CGTGGGGGTGGGAGGAGCCCTCTGTGATTCTTGGCCCTGGCAGTCAGCATCCAG	863	
71	TyrAspLysGlnHisValCysGlySerIleLeuAspProHisTrpValLeuThrAla	90	
864	TACGACAAACAGCAGCTGTGTGGAGGAGCATCTTGGACCCCTCTGGTCTCTCAGGCA	923	
91	AlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySerAsp	110	
924	GCCACATGCTTCAGAAACATACCAGATGTGTCACTGGAAGTGGCGGAGGCTCAGAC	983	
111	LysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleIleGluPheAsnPro	130	
984	AAACTGGGCAGTCCCATCCCTGGCTGTGGCAGAGCATCATCATTAATTCACCCCC	1043	
131	MetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPheSer	150	
1044	ATGTACCCCAAAGACAATGATGATCGCCCTCATGAAGCTGCAGTTCACCATCTCTCTCA	1103	
151	GlyThrValArgProIleCysLeuProPhePheAspGluLeuThrProAlaThrPro	170	
1104	GGCAGAGTCAGGCCCATCTGCTGCCCTTCCTTGTATGAGGAGCTCACTCCAGCCACCCCA	1163	
171	LeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyGlyLysMetSerAspIleLeu	190	
1164	CTCTGGATCATTTGATGGGGCTTTACGAGCAGCAATGGAGGGAAGATGTCGACATCATCTG	1223	
191	LeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAlaTyrGln	210	

[illegible]

```
Y 131 MetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPheSer 150
b 1044 ATGTACCCCAAGACAAATGATCGCCTCATGAAGCTGAGTTCCTCCACTCACTTTCTCA 1103
Y 151 GlyThrValArgProIleCysLeuProPhePheAspGluGluLeuThrProAlaThrPro 170
b 1104 GGACAGTCAGGCCCATCTGTCTCCCTTCTTTGATGAGGAGCTCACTCCAGCCACCCCA 1163
Y 171 LeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyGlyLysMetSerAspIleLeu 190
b 1164 CTCTGGATCATTGGATGGGCTTTACGACAGATGGAGGAGATGTCTGACACTG 1223
Y 191 LeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAlaTyrGln 210
b 1224 CTGCAGCGCTCAGTCCAGGTCATTGACAGCACACGGTGCATGCGACAGATGGTACCAG 1283
Y 211 GlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThrCys 230
b 1284 GGGGAGTCAACGAGAAAGATGATGTGACAGGCATCCCGAAGGGGGTGTGGACACCTGC 1343
Y 231 GlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGlyIle 250
b 1344 CAGGTTGACAGTGGTGGGCCCTGATGTACCAATCTGACCAGTGGCATGTGGTGGGATC 1403
Y 251 ValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyValTyrThrLysValSer 270
b 1404 GTTAGCTGGGGCTATGGCTGGGGGGCCGAGCACCCAGGAGTATACACCAAGGTCTCA 1463
Y 271 AlaTyrLeuAsnTrpIleTyrAsnValTrpIleValAlaGluLeu 284
b 1464 GCCTATCTCACTGGATCTACATGTCTGGAAGGCTGAGCTG 1505
```

Search completed: November 23, 2003, 12:22:38
Job time : 334.091 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.
November 23, 2003, 07:53:00 ; Search time 2075.73 Seconds
(without alignments)
3418.993 Million cell updates/sec

US-09-607-745-9
Exact score: 1600
Sequence: 1 MDSKSSQKSRLLLLLVSN.....LNWYVWKAELSRHHHHH 292

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 22781392 seqs, 12152238056 residues
Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Maximum Match 0%
Listing first 45 summaries

Command line parameters: -DEV=xlh
MODEL=framer_pzn.model -DEVP=0
DB=EST -QFMT=fastap -SUFFIX=est -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0
JNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
OCOCALIGN=200 -THR_SCORE=est -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
USER=us09607745 @CGN_1_3596 @runat_21112003_144346_22229 -NCFU=6 -ICPU=3
NO_WMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6
XGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :
EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_nam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rtd:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*

29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1102.5	68.9	975	10	BG288427	BG288427 602388091
2	1050	65.6	2177	11	AK078890	AK078890 Mus muscu
3	1021	63.8	895	13	BUS22841	BUS22841 AGENCOURT
4	968	60.5	1973	11	BC004855	BC004855 Homo sapi
5	895	55.9	689	12	BG966811	BG966811 602834306
6	885	52.5	515	14	CB142902	CB142902 K-EST0196
7	840	52.5	569	12	BM686292	BM686292 UI-E-CR0-
8	816	51.0	659	9	AI924527	AI924527 wh61607.x
9	808	50.5	648	9	AI924182	AI924182 wh53111.x
10	797	49.8	563	12	BM773306	BM773306 K-EST0057
11	797	49.8	572	12	BM746729	BM746729 K-EST0021
12	797	49.8	578	12	BM747250	BM747250 K-EST0021
13	797	49.8	593	12	BM788452	BM788452 K-EST0067
14	797	49.8	596	12	BM772743	BM772743 K-EST0056
15	797	49.8	614	12	BM788163	BM788163 K-EST0067
16	797	49.8	633	12	BM764659	BM764659 K-EST0046
17	797	49.8	691	12	BM763697	BM763697 K-EST0045
18	797	49.8	692	12	BM772754	BM772754 K-EST0056
19	797	49.8	700	12	BM772620	BM772620 K-EST0056
20	783	48.9	548	12	BM746721	BM746721 K-EST0021
21	780	48.8	920	13	BQ226566	BQ226566 AGENCOURT
22	687	42.9	408	12	BM762292	BM762292 K-EST0043
23	657	41.1	584	9	AI597567	AI597567 tr92a03.x
24	656.5	41.0	607	9	AA143746	AA143746 z067a02.x
25	655	40.9	711	10	BE937482	BE937482 RC4-ST027
26	652	40.8	435	10	BF811390	BF811390 CM2-C1017
27	648.5	40.5	644	12	BM791640	BM791640 K-EST0071
28	648	40.5	481	9	AA143761	AA143761 z067a02.x
29	637	39.8	721	13	BUG00097	BUG00097 AGENCOURT
30	618	38.6	425	10	AW908659	AW908659 uf81c03.y
31	609	38.1	956	13	BUS23218	BUS23218 AGENCOURT
32	603	37.7	542	10	AW913122	AW913122 uf49b12.y
33	589	36.8	569	9	AW751450	AW751450 MR1-CT005
34	558	34.9	702	12	BM695610	BM695610 UI-E-C01-
35	551	34.4	498	9	AI158537	AI158537 wd25a02.x
36	538.5	33.7	1835	11	BC048135	BC048135 Danilo fer
37	531.5	33.2	1172	14	CD508404	CD508404 CDA90-Fil
38	519	32.4	2895	11	AK029714	AK029714 Mus muscu
39	514.5	32.2	1173	14	CD508765	CD508765 CDA92-F09
40	506	31.6	1749	11	BC035623	BC035623 Homo sapi
41	502	31.4	586	9	AW613143	AW613143 hh42c04.x
42	500	31.2	683	13	BUS29098	BUS29098 604163681
43	493	30.8	2055	11	BC028065	BC028065 Homo sapi
44	491	30.7	3035	11	AK004939	AK004939 Mus muscu
45	481	30.1	816	13	BUS397449	BUS397449 603536894

ALIGNMENTS

RESULT 1
BG288427
LOCUS 602388091F1 NIH_MGC_93 Homo sapiens cDNA clone IMAGE:4516990 5',
DEFINITION 975 bp mRNA linear EST 21-FEB-2001
mRNA sequence.
ACCESSION BG288427
VERSION BG288427.1 GI:13043459
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 975)

5', mRNA sequence.

CESSION
CB142902
VERSION
CB142902.1 GI:28119586
EST.

SOURCE

Homosapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 515)
Kim.N.S., Hahn.Y., Oh.J.H., Lee.J.Y., Ahn.H.Y., Chu.M.Y., Kim.M.R.,
Oh.K.J., Cheong.J.E., Sohn.H.Y., Kim.J.M., Park.H.S., Kim.S. and
Kim.Y.S.

21C Frontier Korean EST Project 2001

Unpublished

Contact: Kim YS

Genome Research Center

Korea Research Institute of Bioscience & Biotechnology

52 Boeun-dong Yuseong-gu, Daejeon 305-333, South Korea

Tel: +82-42-860-4470

Fax: +82-42-860-4409

Email: yongsung@mail.kribb.re.kr

Plate: 20 row: D column: 07

High quality sequence stop: 515.

Location/Qualifiers

FEATURES

source

1..515

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="L118NU354s1-20-D07"

/sex="M"

/tissue_type="Liver"

/cell_type="Polygonal"

/cell_line="SNU-354"

/lab_host="Top10P"

/clone_lib="L118NU354s1"

/note="Organ: Liver; Vector: pCNS-D2; Site 1: EcoRI;

Site 2: NotI; The poly (A)+ RNA was dephosphorylated with

bacterial alkaline phosphatase (BAP) and then decapped

with tobacco acid pyrophosphatase (TAP). The decapped

intact mRNA was ligated with DNA-RNA linker including

EcoRI site by treatment of T4 RNA ligase and the first

strand cDNA was synthesized from oligo dT-selected mRNA by

priming with dN-tailed vector. The dT-tailed vector was

adjusted to have about 60nt. The cDNA vector was

circularized with E. coli DNA ligase after digestion of

EcoRI which site is also included in vector. An RNA strand

converted to a DNA strand by Okayama-Berg method. The

obtained cDNA vectors were used for transformation of

competent cells E. coli Top10P' by electroporation method.

The cDNA libraries constructed by this method are

full-length enriched cDNA library. After analyzing and

sequencing about 2,000 - 3,000 colonies in original cDNA

library, the abundant cDNAs were selected and amplified by

PCR reaction using vector region primer including T7

promotor as 5' primer and N(dT)14 as 3' primer. The PCR

products were used as template for synthesis of

biotinylated single stranded RNA by in vitro transcription

reaction. The synthesized RNA probes were hybridized with

antisense single stranded cDNAs prepared from original

library and incubated with avidin-gel. After removing

DNA-RNA hybrids by centrifuge, the subtracted cDNA

libraries were constructed by transformation of the

remaining DNA into competent cells E. coli Top10P' with

electroporation method."

BASE COUNT

116 a 139 c 150 g 110 t

Alignment Scores:

ed. No.: 3.31e-90 Length: 515
Score: 885.00 Matches: 160
Percent Similarity: 100.00% Conservative: 0
st Local Similarity: 100.00% Mismatches: 0
Query Match: 55.31% Indels: 0

DB: 14 Gaps: 0

US-09-607-745-9 (1-292) x CB142902 (1-515)

QY 125 IleIleGluPheAsnProMetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGln 144

Db 2 ATCATTTGAATTCACCCCATGTACCCCAAGACAATGATGCCCTCATGAAGCTGCAG 61

QY 145 PheProLeuThrPheSerGlyThrValArgProLysCysLeuProPhePheAspGluGln 164

Db 62 TTCACACTCATTCTCAGGCACAGTCAGGCCCATCTGTCCTCTCTTGTATGAGGAG 121

QY 165 LeuThrProAlaThrProLeuThrPheIleGlyThrGlyPheThrLysGlnAsnGlyGly 184

Db 122 CTCACCTCCAGCCACCCCACTCTGGATCATTTGGATGGGGTTTACGAGCAGATGGAGG 181

QY 185 LysMetSerAspIleLeuGlnLysValGlnValIleAspSerThrArgCysAsn 204

Db 182 AAGATGTCTGACATCTCTGCAGCGCTCAGTCCAGGTCAATGACAGCACACGGTGCAT 241

QY 205 AlaAspAspAlaTyrGlnGlyGlnValThrGluLysMetMetCysAlaGlyTleProGlu 224

Db 242 GCAGACGATGCGTACCGGGGAGTCCCGGAGAGATGTGTGCGAGGCGATCCCGGAA 301

QY 225 GlyGlyValAspThrCysGlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGln 244

Db 302 GGGGGTGTGGACACCTGCCAGGTGACAGTGGTGGCCCCCTGATGTACCAATCTGACCCAG 361

QY 245 TrpHisValValGlyIleValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGly 264

Db 362 TGGCATGTGGTGGGCGATCGTTAGTTGGGGCTATGCTGCGGGGGCCGAGCACCCAGGA 421

QY 265 ValTyrThrLysValSerAlaTyrIleuAsnTrpIleTyrAsnValTrpLysAlaGluLeu 284

Db 422 GTATACACCAAGGCTCAGCCTAUCTCACTGGATCTACAAATGTCTGAAGCTGAGCTG 481

BM686292 569 bp mRNA linear EST 28-FEB-2002

UI-E-CR0-adk-g-09-0-UI-r1 UI-E-CR0 Homo sapiens cDNA clone

UI-E-CR0-adk-g-09-0-UI 5', mRNA sequence.

BM686292

BM686292.1 GI:18999550

EST.

Source

Homo sapiens (human)

Organism

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 569)

Bonaldi,M.F., Lennon,G. and Soares,M.B.

Normalization and subtraction: two approaches to facilitate gene

discovery

Genome Res. 6 (9), 791-806 (1996)

97044477

8889548

Contact: Soares, MB

Coordinated Laboratory for Computational Genomics

University of Iowa

375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA

Tel: 319 335 8250

Fax: 319 335 9565

Email: bento-soares@uiowa.edu

Tissue Procurement: Dr. Gregg Hageman

cDNA Library preparation: Dr. M. Bento Soares, University of Iowa

CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa

CDNA Sequencing by: Dr. M. Bento Soares, University of Iowa

Clone Distribution: Researchers may obtain clones from Research

Genetics (www.resgen.com).

Seq primer: M13 Reverse.

Location/Qualifiers

1..569

/organism="Homo sapiens"

/mol_type="mRNA"

FEATURES

source


```

/db_xref="taxon:9606"
/clone="UI-E-CRO-adj-g-09-0-UI"
/tissue_type="eye anterior segment"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/clone_lib="UI-E-CRO"
/notes="Organ: eye; Vector: pT73-Pac (Pharmacia) with a modified polylinker; Site 1: EcoR I; Site 2: Not I; UI-E-CRO is a cDNA library containing the following tissue(s): eye anterior segment. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT73-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (drr)18 tail. The sequence tag for this library is AATGCCGCAT. This library was created for the program, Gene Discovery in the Visual System, supported by National Eye Institute (NRI)."

```

SE COUNT	124 a	169 c	156 g	120 t
Alignment Scores:				
ed. No.:	5.04e-85		Length:	569
ore:	840.00		Matches:	156
Percent Similarity:	94.15%		Conservative:	5
st Local Similarity:	91.23%		Mismatches:	8
ery Match:	52.50%		Indels:	2
	12		Gaps:	1

-09-607-745-9 (1-292) x BM686292 (1-569)

```

41 AlaLeuAlaAProPheAsp-----AspAspLysIleValGlyGlyTyrAlaLeu 58
      ::::: ::::: :::::
55 GCACGTCTGTGCTGTGGGAAACCTGAAGACCCCTGTGTGGTGGGGAGGAGGC 114
      ::::: ::::: :::::
59 AspValAspSerTrpProTrpGlnValSerIleGlnTyrAspLysGlnHisValCysGly 78
      ::::: ::::: :::::
115 TCTGTGGATTCTTGCGCTTGGCAGTGCAGTACGACAAACAGCACGCTCTGTGGA 174
      ::::: ::::: :::::
179 GlySerIleLeuAspProHisTrpValLeuThrAlaAlaHisCysPheArgLysHisThr 98
      ::::: ::::: :::::
175 GGGAGCATCTTGGACCCCACTGGGGTCTTACGGAGCCCACTGCTTCAGAAACATACC 234
      ::::: ::::: :::::
99 AspValPheAsnTrpLysValArgAlaGlySerAspLysLeuGlySerPheProSerLeu 118
      ::::: ::::: :::::
235 GATGTGTTCAACTGGAGGTGGGGCAGGCTCAGACAAACTGGGGCAGCTCCCATCCCTG 294
      ::::: ::::: :::::
119 AlaValAlaLysIleIleIleGluPheAsnProMetTyrProLysAspAsnAspIle 138
      ::::: ::::: :::::
295 GTGTGGCCCAAGATCATCATATTGAATTCAACCCCATGTACCCAAAGACATGACATC 354
      ::::: ::::: :::::
139 AlaLeuMetLysLeuGlnPheProLeuThrPheSerGlyThrValArgProIleCysLeu 158
      ::::: ::::: :::::
355 GCCTCATGAAGCTCAGTCCCACTCACTTTCAGGCACAGTCAGGCCCATCTGCTG 414
      ::::: ::::: :::::
159 ProphePheAspGluGluLeuThrProAlaThrProLeuTrpIleIleGlyTrpGlyPhe 178
      ::::: ::::: :::::
415 CCTTCTTTGATGAGGAGCTCACTCCAGCCACCACCTCTGGATCATTTGATGGGGCTTT 474
      ::::: ::::: :::::
179 ThrLysGlnAsnGlyGlyLysMetSerAspIleLeuLeuGlnAlaSerValGlnValIle 198
      ::::: ::::: :::::
475 ACGAAGACAGAAATGGAGGGAAGATGCTGACATACTGTCAGGGCGCTCACTCCAGGTCA 534
      ::::: ::::: :::::
199 AspSerThrArgCysAsnAlaAspAlaTyr 209
      ::::: ::::: :::::
535 GCAGCACACCGGTGCAATGCAGACGATCGGTAC 567

```

SULT 8
024527/c

LOCUS	AI924527	659 bp	mrna	linear	EST 08-MAR-2000
DEFINITION	wn1c07.x1 NCI CGAP lnl9 Homo sapiens cDNA clone IMAGE:2449932 3', similar to TR:060235 060235 AIRWAY TRYPSIN-LIKE PROTEASE. ;, mRNA sequence.				
ACCESSION	AI924527				
VERSION	AI924527.1	GI:5660491			
KEYWORDS	EST.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
REFERENCE	1 (bases 1 to 659)				
AUTHORS	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap .				
TITLE	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index				
JOURNAL	Unpublished				
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgapbs-r@mail.nih.gov Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D. cDNA Library Prepared by: M. Bento Soares, Ph.D. cDNA Library Arrayed by: Greg Lennon, Ph.D. DNA Sequencing by: Washington University Genome Sequencing Center Clone Distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html Insert Length: 1830 Std Error: 0.00 Seq primer: -40UP from Gibco High quality sequence stop: 450.				

FEATURES
source

```

/organism="Homo sapiens"
/mol_type="rRNA"
/db_xref="taxon:9606"
/clone="IMAGE:244932"
/tissue_type="squamous c
differentiated (4 pooled
metastatic)"
/dev_stage="adult"
/lab_host="DH10B (phage-
/clonelib="NCI CGAP Lul
/notes="Organ: lung; Vect
modified polylinker; 1st
pooled lung tumor tissue
- cligo (dr) primer. Doub
Eco RI adaptors (Pharmac
cloned into the Not I a
pRT73 vector. Library w
normalization. Library c
Fatima, Ronaldo. "
```

FALLIMA BONARDO.			
BASE COUNT	148 a	176 c	178 g
157 t			

Alignment Scores:	
Pred. No.:	3.38e-82
Score:	816.00
Length:	659
Matches:	147
Percent Similarity:	100.00%
Best Local Similarity:	100.00%
Mismatches:	0
Query Match:	51.00%
Indels:	0
DB:	9
Gaps:	0

US-09-607-745-9 (1-292) x AI924527 (1-659)

Qy	135	AspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPheSerGlyThrValArg	154
Db	659	GACAATGACATCGCCCTCATGAAGTCGAGTCCCACTCACTTCTCAGGCACAGTCAGG	600
Qy	155	ProIleCysLeuProPhePheAspGluGluLeuThrProAlaThrProLeuTrpIleIle	174
Db	599	CCCATCTGCTGCCCCCTCTTTGAGGAGGAGCTCACGCCAGCCACCCACTCTGGATCAAT	540
Qy	175	GlyTrpGlyPheThrLysGlnAsnGlyGlyLysMetSerAspIleLeuGlnAlaSer	194

```

b 539 GGATGGGGCTTTACGAGCAGCAATGGAGGAGATGCTGACATACCTGCTGCAGGGCTCA 480
Y 195 valcInValIleAspSerThrArgCysAsnAlaAspAlaTyrGlnGlyGluValThr 214
b 479 GTTCAGGTCATTCACACACACACGGTGAATGCAGACGATCGTACCAAGGGGGAAGTCACC 420
Y 215 GluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThrCysGlnGlyAspSer 234
b 419 GAGAGATGATGTGTGCAGGATCCCGGAAGGGGTGTGCACACCTGCCAGGGTGACAGT 360
Y 235 GlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGlyIleValSerTrpGly 254
b 359 GGTGGGGCCCTGATGATACCAATCTGACCACTGGCATGTGTGGGCATCGTATGCGGC 300
Y 255 TyrGlyCysGlyGlyProSerThrProGlyValTyrThrLysValSerAlaTyrLeuAsn 274
b 299 TATGGTGTGGGGGGCCCGAGCACCAGGAGGTATACACCAAGGTCTCAGCCTATCTCAAC 240
Y 275 TrpIleTyrAsnValTrpLys 281
b 239 TGGATCTACAAATGCTCGGAAG 219

RESULT 9
1924182/c
LOCUS
DEFINITION
  648 bp mRNA linear EST 02-SEP-1999
  wn53f11.x1 NCI CGAP Jul19 Homo sapiens cDNA clone IMAGE:2449197 3'
  similar to TR:060235 O60235 AIRWAY TRYPSIN-LIKE PROTEASE. ;, mRNA
  sequence.
  AI924182
  AI924182.1 GI:5660146
  EST.
  Homo sapiens (human)
  SOURCE
  Homo sapiens
  ORGANISM
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  1 (bases 1 to 648)
  NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
  Tumor Gene Index
  Unpublished
  Contact: Robert Strausberg, Ph.D.
  Email: cgapsb@mail.nih.gov
  Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
  Emmert-Buck, M.D., Ph.D.
  cDNA Library Preparation: M. Bento Soares, Ph.D.
  DNA Sequencing Arrayed by: Greg Lennon, Ph.D.
  Clone Distribution: NCI-CGAP clone distribution information can be
  found through the I.M.A.G.E. Consortium/LLNL at:
  www-bio.llnl.gov/bbrp/image/image.html
  Seq primer: -40UP from Gibco
  High quality sequence stop: 455.
  Location/Qualifiers
    1..648
    /organism="Homo sapiens"
    /mol_type="mRNA"
    /db_xref="taxon:9606"
    /clone IMAGE:2449197"
    /tissue_type="squamous cell carcinoma, poorly
    differentiated (4 pooled tumors, including primary and
    metastatic)"
    /dev_stage="adult"
    /lab_host="PH10B (phage-resistant)"
    /clone_lib="NCI_CGAP_Jul19"
    /note="Organ: lung; Vector: pT7T3D-Pac (Pharmacia) with a
    modified polylinker; 1st strand cDNA was prepared from
    pooled lung tumor tissue, and was then primed with a Not I
    - oligo(dT) primer. Double-stranded cDNA was ligated to
    Eco RI adaptors (Pharmacia), digested with Not I and
    cloned into the Not I and Eco RI sites of the modified
    pT7T3 vector. Library went through one round of
    normalization. Library constructed by Bento Soares and M.
    Fatima Bonaldo."

```

```

BASE COUNT      144 a      174 c      180 g      150 t
ORIGIN

Alignment Scores:
Pred. No.:      2,68e-81      Length:      648
Score:          808.00      Matches:    146
Percent Similarity: 99.32%      Conservative: 0
Best Local Similarity: 99.32%      Mismatches: 1
Query Match:     50.50%      Indels:    0
DB:              9      Gaps:      0

```

US-09-607-745-9 (1-292) x AI924182 (1-648)

```

QY 135 AspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPheSerGlyThrValArg 154
Db 648 GACAATGACATCGCCTCATGAAGCTGCAGTTCCTCCACTCACCTTCTCAGGCACAGTCAGG 589
QY 155 ProIleCysLeuProPheAspGluGluLeuThrProAlaThrProLeuTrpIleIle 174
Db 588 CCCATCTGCTGCCCTTCTTTGATGAGGAGCTCACTCAAGCCACCCCACTCTGGATCAAT 529
QY 175 GlyTrpGlyPheThrLysGlnAsnGlyLysMetSerAspIleLeuGlnAlaSer 194
Db 528 GGATGGGGCTTTTACGAAGCAGAAATGGAGGAAGATGTCTGACATACTGCTGAGGGGTCA 469
QY 195 ValGlnValIleAspSerThrArgCysAsnAlaAspAlaTyrGlnGlyGluValThr 214
Db 468 GTCCAGGTCAATGACAGCACACAGCGTGCAATGCAGACGATCGGTACAGGGGGAAGTCACC 409
QY 215 GluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThrCysGlnGlyAspSer 234
Db 408 GAGAAAGATGATGTGTGCAGGCATCCCGGAAGGGGTGTGGACACCTGCCAGGGTGACACT 349
QY 235 GlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGlyIleValSerTrpGly 254
Db 348 GTGGGGCCCTGATGTACCAATCTACCAGTGGCATGTGGGGCATCGTTAGCTGGGGC 289
QY 255 TyrGlyCysGlyGlyProSerThrProGlyValTyrThrLysValSerAlaTyrLeuAsn 274
Db 288 TATGGCTGGGGGGCCCGAGCACCCAGAGATATACACCAAGGTCTCAGCCTATCTCAAC 229
QY 275 TrpIleTyrAsnValTrpLys 281
Db 228 TGGATCTACAAATGCTCGGAAG 208

```

```

RESULT 10
BM773306
LOCUS
DEFINITION
  K-EST0057738 S3SNU16s1 Homo sapiens cDNA clone S3SNU16s1-17-C08 5',
  mRNA sequence.
  BM773306
  BM773306.1 GI:19102921
  EST.
  KEYWORDS
  SOURCE
  Homo sapiens (human)
  ORGANISM
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  1 (bases 1 to 563)
  Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
  Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and
  Kim,Y.S.
  21C Frontier Korean EST Project 2001
  Unpublished
  Contact: Kim YS
  Genome Research Center
  Korea Research Institute of Bioscience & Biotechnology
  52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea
  Tel: +82-42-860-4470
  Fax: +82-42-860-4409
  Email: yongsung@mail.kribb.re.kr
  Plate: 17 row: C column: 08
  High quality sequence stop: 563.
  Location/Qualifiers

```

FEATURES

source

```

1. .563
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="S3SNU16s1-17-C08"
/sex="F"
/tissue_type="Ascites"
/cell_type="Lymphoblast-like"
/lab_host="SNU-16"
/clone_lib="S3SNU16s1"
/notes="Organ: Stomach; Vector: pTZ18RPl; Site 1: EcoRI;
Site 2: NotI; The poly (A) + RNA was dephosphorylated with
bacterial alkaline phosphatase (BAP) and then decapped
with tabacco acid pyrophosphatase (TAP). The decapped
intact mRNA was ligated with DNA-RNA linker including EcoR
I site by treatment of T4 RNA ligase and the first strand
cDNA was synthesized from oligo dt-selected mRNA by
priming with dt-tailed vector. The dt-tailed vector was
adjusted to have about 60nt. The cDNA vector was
circularized with E. coli DNA ligase after digestion of
EcoRI which site is also included in vector. An RNA strand
obtained cDNA vectors were used for transfection of
competent cells E. coli Top10F' by electroporation method.
The cDNA libraries constructed by this method are
full-length enriched cDNA library. After analyzing and
sequencing about 2,000 ~ 3,000 colonies in original cDNA
library, the abundant cDNAs were selected and amplified by
PCR reaction using vector region primer including T7
promotor as 5' primer and N(dt)14 as 3' primer. The PCR
products were used as template for synthesis of
biotinylated single stranded RNA by in vitro transcription
reaction. The synthesized RNA probes were hybridized with
antisense single stranded cDNAs prepared from original
library and incubated with avidin-gel. After removing
DNA-RNA hybrids by centrifuge, the substracted cDNA
libraries were constructed by transfection of the
remaining DNA into competent cells E. coli Top10F' with
electroporation method."
BASE COUNT      121 a 172 c 148 g 122 t
ORIGIN

Alignment Scores:
red. No.:      3,84e-80      Length:      563
core:          797.00      Matches:     151
percent Similarity: 75.50%      Conservative: 0
nearest Local Similarity: 75.50%      Mismatches:  0
query Match:     49.81%      Indels:      49
B:              12      Gaps:         1

us-09-607-745-9 (1-292) x BM773306 (1-563)

86 TrpValLeuThrAlaAlaHisCysPheArgLysHisThrAspValPheAsnTrpLysVal 105
2 TGGGTCTCAAGGAGCCCTGCTTCAGAACATACCGATGTGTTCACATGGAGGTG 61
106 ArgAlaGlySerAspLysLeuGlySerPheProSerLeuAlaValAlaLysLeilelle 125
62 CGGGCAGGCTCAGACAAACTGGGCAGCTTCCCATCCCTGGCTGTGCCAAGATCATCATC 121
126 IleGluPheAsnProMetTrpProLysAspAsnAspIleAlaLeuMetLysLeuGlnPhe 145
122 ATTGATTCACACCCCATGACCCCAAGACATGATCGCCTCATGAAGCTGCAGTTC 181
146 ProLeuThrPheSerGlyThrValArgProIleCysLeuProPhePheAspGluGluLeu 165
182 CCACCTCACTTCTCAGGCACATCAGGCCCATCTGTCTGCCCTTCTTGTATGAGAGCTC 241
166 ThrProAlaThrProLeuTrpIlelleGlyTrpGlyPheThrLysGlnAsnGlyGlyLys 185
242 ACTCCAGCACCCCACTCTGGATCATTCGATGGGCTTTACGACGAGATGA----- 295

```

```

QY 186 MetSerAspIleLeuLeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAla 205
Db 295 -----
QY 206 AspAspAlaTrpGlnGlyGluValThrGluLysMetMetCysAlaGlyIleProGluGly 225
Db 295 -----
QY 226 GlyValAspThrCysGlnGly-AspSerGlyGlyProLeuMetTrpGlnSerAspGlnTr 245
Db 296 -----CGGTGACAGTGGTGGGCCCTGATGATACCAATCTGACCAGTG 337
QY 245 pHisValValGlyIleValSerTrpGlyTyrrGlyCysGlyGlyProSerThrProGlyVa 265
Db 338 GCATGTGTGTGGGCATCGTTAGTTGGGCTATGGCTCGGGGGCCCGAGCACCCAGGAGT 397
QY 265 lTyrrThrLysValSerAlaTyrrLeuAsnTrpIleTyrrAsnValTrpLysAlaGluLeu 284
Db 398 ATACACCAAGGCTCTAGCCTATCTCACTGGATCTCAATGTCTGAAGGCTGAGCTG 455

RESULT 11
BM746729
LOCUS
DEFINITION
X-EST0021071 S3SNU16 Homo sapiens cDNA clone S3SNU16-19-D01 5',
mRNA sequence.
ACCESSION
BM746729
VERSION
BM746729.1 GI:19076313
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and
Kim,Y.S.
21C Frontier Korean EST Project 2001
JOURNAL
Unpublished
COMMENT
Contact: Kim YS
Genome Research Center
Korea Research Institute of Bioscience & Biotechnology
52 Boeun-dong Yuseong-gu, Daejeon 305-333, South Korea
Tel: +82-42-860-4470
Fax: +82-42-860-4409
Email: yongsung@mail.kribb.re.kr
Plate: 19 row: D column: 01
High quality sequence stop: 572.
FEATURES
Location/Qualifiers
1..572
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="S3SNU16-19-D01"
/sex="F"
/tissue_type="Ascites"
/cell_type="Lymphoblast-like"
/lab_host="SNU-16"
/clone_lib="S3SNU16"
/notes="Organ: Stomach; Vector: pTZ18RPl; Site 1: EcoRI;
Site 2: NotI; The poly (A) + RNA was dephosphorylated with
bacterial alkaline phosphatase (BAP) and then decapped
with tabacco acid pyrophosphatase (TAP). The decapped
intact mRNA was ligated with DNA-RNA linker including EcoR
I site by treatment of T4 RNA ligase and the first strand
cDNA was synthesized from oligo dt-selected mRNA by
priming with dt-tailed vector. The dt-tailed vector was
adjusted to have about 60nt. The cDNA vector was
circularized with E. coli DNA ligase after digestion of
EcoRI which site is also included in vector. An RNA strand
obtained cDNA vectors were used for transfection of
competent cells E. coli Top10F' by electroporation method."

```


122 ATTGAATCAACCCCATGTATCCCAAGACAAATGACATGCGCCCTCATGAAGCTGCAGTTC 181
 146 ProLeuThrPheSerGlyThrValArgProIleCysLeuProPhePheAspGluGluLeu 165
 182 CCACTCACATTCTCAGGCACAGTCAGGCCCATCTGTCTGCCCTTCTTTGATGAGGAGCTC 241
 166 ThrProAlaThrProLeuThrPheIleGlyTyrGlyPheThrLysGlnAsnGlyGlyLys 185
 242 ACTCCAGGCCACCCCACTCTGGATCATTTGGATGGGGCTTTACGAGACGAATGGA----- 295
 186 MetSerAspIleLeuLeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAla 205
 295 ----- 295
 206 AspAspAlaTyrGlnGlyGluValThrGluLysMetMetCysAlaGlyIleProGluGly 225
 295 ----- 295
 226 GlyValAspThrCysGlnGly-AspSerGlyGlyProLeuMetTyrGlnSerAspGlnTr 245
 296 -----GGGTGACAGTGTGGGGCCCTTGATGATCAATCTGACCAAGT 337
 245 pHisValValGlyIleValSerTyrGlyTyrGlyCysGlyGlyProSerThrProGlyVa 265
 338 GCATGTGTGGGCATGTTAGTTGGGCTATGCTGGGGGGCCCGAGCACCCAGGAGT 397
 265 lTyrThrLysValSerAlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu 284
 398 ATACACCAAGGTCTCAGCCTATCTCACTGGATGATCAATGATGCTGGAAGGCTGAGCTG 455

RESULT 13
 M788452
 LOCUS
 DEFINITION
 K-EST0056963 S3SNUI6s1 Homo sapiens cdna clone S3SNUI6s1-36-All 5',
 mRNA sequence.
 BM788452
 BM788452.1 GI:19136684
 EST.
 Homo sapiens (human)
 Homo sapiens
 Mammalia; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 593)
 Kim.N.S., Hahn.Y., Oh.J.H., Lee.J.Y., Ahn.H.Y., Chu.M.Y., Kim.M.R.,
 Oh.K.J., Cheong.J.E., Sohn.H.Y., Kim.J.M., Park.H.S., Kim.S. and
 Kim.Y.S.
 21C Frontier Korean EST Project 2001
 Unpublished
 Contact: Kim YS
 Genome Research Center
 Korea Research Institute of Bioscience & Biotechnology
 52 Boeun-dong Yuseong-gu, Daejeon 305-333, South Korea
 Tel: +82-42-860-4470
 Fax: +82-42-860-4409
 Email: yongsung@mail.kribb.re.kr
 Plate: 36 row: A column: 11
 High quality sequence stop: 593.
 Location/Qualifiers
 1..593
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="S3SNUI6-36-All"
 /sex="F"
 /tissue_type="Ascites"
 /cell_type="Lymphoblast-like"
 /cell_line="SNU-16"
 /lab_host="Top10P"
 /clone_lib="S3SNUI6"
 /note="Organ: Stomach; Vector: pTZ18RP1; Site: 1: EcoRI;
 Site 2: NotI; The poly (A)+ RNA was dephosphorylated with
 bacterial alkaline phosphatase (BAP) and then decapped
 with tobacco acid pyrophosphatase (TAP). The decapped

BM788452 593 bp mRNA linear EST 05-MAR-2002
 K-EST0056963 S3SNUI6 Homo sapiens cdna clone S3SNUI6s1-36-All 5',
 mRNA sequence.
 BM788452
 BM788452.1 GI:19136684
 EST.
 Homo sapiens (human)
 Homo sapiens
 Mammalia; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 593)
 Kim.N.S., Hahn.Y., Oh.J.H., Lee.J.Y., Ahn.H.Y., Chu.M.Y., Kim.M.R.,
 Oh.K.J., Cheong.J.E., Sohn.H.Y., Kim.J.M., Park.H.S., Kim.S. and
 Kim.Y.S.
 21C Frontier Korean EST Project 2001
 Unpublished
 Contact: Kim YS
 Genome Research Center
 Korea Research Institute of Bioscience & Biotechnology
 52 Boeun-dong Yuseong-gu, Daejeon 305-333, South Korea
 Tel: +82-42-860-4470
 Fax: +82-42-860-4409
 Email: yongsung@mail.kribb.re.kr
 Plate: 36 row: A column: 11
 High quality sequence stop: 593.
 Location/Qualifiers
 1..593
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="S3SNUI6-36-All"
 /sex="F"
 /tissue_type="Ascites"
 /cell_type="Lymphoblast-like"
 /cell_line="SNU-16"
 /lab_host="Top10P"
 /clone_lib="S3SNUI6"
 /note="Organ: Stomach; Vector: pTZ18RP1; Site: 1: EcoRI;
 Site 2: NotI; The poly (A)+ RNA was dephosphorylated with
 bacterial alkaline phosphatase (BAP) and then decapped
 with tobacco acid pyrophosphatase (TAP). The decapped

BM788452 593 bp mRNA linear EST 05-MAR-2002
 K-EST0056963 S3SNUI6 Homo sapiens cdna clone S3SNUI6s1-36-All 5',
 mRNA sequence.
 BM788452
 BM788452.1 GI:19136684
 EST.
 Homo sapiens (human)
 Homo sapiens
 Mammalia; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 593)
 Kim.N.S., Hahn.Y., Oh.J.H., Lee.J.Y., Ahn.H.Y., Chu.M.Y., Kim.M.R.,
 Oh.K.J., Cheong.J.E., Sohn.H.Y., Kim.J.M., Park.H.S., Kim.S. and
 Kim.Y.S.
 21C Frontier Korean EST Project 2001
 Unpublished
 Contact: Kim YS
 Genome Research Center
 Korea Research Institute of Bioscience & Biotechnology
 52 Boeun-dong Yuseong-gu, Daejeon 305-333, South Korea
 Tel: +82-42-860-4470
 Fax: +82-42-860-4409
 Email: yongsung@mail.kribb.re.kr
 Plate: 36 row: A column: 11
 High quality sequence stop: 593.
 Location/Qualifiers
 1..593
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="S3SNUI6-36-All"
 /sex="F"
 /tissue_type="Ascites"
 /cell_type="Lymphoblast-like"
 /cell_line="SNU-16"
 /lab_host="Top10P"
 /clone_lib="S3SNUI6"
 /note="Organ: Stomach; Vector: pTZ18RP1; Site: 1: EcoRI;
 Site 2: NotI; The poly (A)+ RNA was dephosphorylated with
 bacterial alkaline phosphatase (BAP) and then decapped
 with tobacco acid pyrophosphatase (TAP). The decapped

BM788452 593 bp mRNA linear EST 05-MAR-2002
 K-EST0056963 S3SNUI6 Homo sapiens cdna clone S3SNUI6s1-36-All 5',
 mRNA sequence.
 BM788452
 BM788452.1 GI:19136684
 EST.
 Homo sapiens (human)
 Homo sapiens
 Mammalia; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 593)
 Kim.N.S., Hahn.Y., Oh.J.H., Lee.J.Y., Ahn.H.Y., Chu.M.Y., Kim.M.R.,
 Oh.K.J., Cheong.J.E., Sohn.H.Y., Kim.J.M., Park.H.S., Kim.S. and
 Kim.Y.S.
 21C Frontier Korean EST Project 2001
 Unpublished
 Contact: Kim YS
 Genome Research Center
 Korea Research Institute of Bioscience & Biotechnology
 52 Boeun-dong Yuseong-gu, Daejeon 305-333, South Korea
 Tel: +82-42-860-4470
 Fax: +82-42-860-4409
 Email: yongsung@mail.kribb.re.kr
 Plate: 36 row: A column: 11
 High quality sequence stop: 593.
 Location/Qualifiers
 1..593
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="S3SNUI6-36-All"
 /sex="F"
 /tissue_type="Ascites"
 /cell_type="Lymphoblast-like"
 /cell_line="SNU-16"
 /lab_host="Top10P"
 /clone_lib="S3SNUI6"
 /note="Organ: Stomach; Vector: pTZ18RP1; Site: 1: EcoRI;
 Site 2: NotI; The poly (A)+ RNA was dephosphorylated with
 bacterial alkaline phosphatase (BAP) and then decapped
 with tobacco acid pyrophosphatase (TAP). The decapped

BM788452 593 bp mRNA linear EST 05-MAR-2002
 K-EST0056963 S3SNUI6 Homo sapiens cdna clone S3SNUI6s1-36-All 5',
 mRNA sequence.
 BM788452
 BM788452.1 GI:19136684
 EST.
 Homo sapiens (human)
 Homo sapiens
 Mammalia; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 593)
 Kim.N.S., Hahn.Y., Oh.J.H., Lee.J.Y., Ahn.H.Y., Chu.M.Y., Kim.M.R.,
 Oh.K.J., Cheong.J.E., Sohn.H.Y., Kim.J.M., Park.H.S., Kim.S. and
 Kim.Y.S.
 21C Frontier Korean EST Project 2001
 Unpublished
 Contact: Kim YS
 Genome Research Center
 Korea Research Institute of Bioscience & Biotechnology
 52 Boeun-dong Yuseong-gu, Daejeon 305-333, South Korea
 Tel: +82-42-860-4470
 Fax: +82-42-860-4409
 Email: yongsung@mail.kribb.re.kr
 Plate: 36 row: A column: 11
 High quality sequence stop: 593.
 Location/Qualifiers
 1..593
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="S3SNUI6-36-All"
 /sex="F"
 /tissue_type="Ascites"
 /cell_type="Lymphoblast-like"
 /cell_line="SNU-16"
 /lab_host="Top10P"
 /clone_lib="S3SNUI6"
 /note="Organ: Stomach; Vector: pTZ18RP1; Site: 1: EcoRI;
 Site 2: NotI; The poly (A)+ RNA was dephosphorylated with
 bacterial alkaline phosphatase (BAP) and then decapped
 with tobacco acid pyrophosphatase (TAP). The decapped

BM788452 593 bp mRNA linear EST 05-MAR-2002
 K-EST0056963 S3SNUI6 Homo sapiens cdna clone S3SNUI6s1-36-All 5',
 mRNA sequence.
 BM788452
 BM788452.1 GI:19136684
 EST.
 Homo sapiens (human)
 Homo sapiens
 Mammalia; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 593)
 Kim.N.S., Hahn.Y., Oh.J.H., Lee.J.Y., Ahn.H.Y., Chu.M.Y., Kim.M.R.,
 Oh.K.J., Cheong.J.E., Sohn.H.Y., Kim.J.M., Park.H.S., Kim.S. and
 Kim.Y.S.
 21C Frontier Korean EST Project 2001
 Unpublished
 Contact: Kim YS
 Genome Research Center
 Korea Research Institute of Bioscience & Biotechnology
 52 Boeun-dong Yuseong-gu, Daejeon 305-333, South Korea
 Tel: +82-42-860-4470
 Fax: +82-42-860-4409
 Email: yongsung@mail.kribb.re.kr
 Plate: 36 row: A column: 11
 High quality sequence stop: 593.
 Location/Qualifiers
 1..593
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="S3SNUI6-36-All"
 /sex="F"
 /tissue_type="Ascites"
 /cell_type="Lymphoblast-like"
 /cell_line="SNU-16"
 /lab_host="Top10P"
 /clone_lib="S3SNUI6"
 /note="Organ: Stomach; Vector: pTZ18RP1; Site: 1: EcoRI;
 Site 2: NotI; The poly (A)+ RNA was dephosphorylated with
 bacterial alkaline phosphatase (BAP) and then decapped
 with tobacco acid pyrophosphatase (TAP). The decapped

BM788452 593 bp mRNA linear EST 05-MAR-2002
 K-EST0056963 S3SNUI6 Homo sapiens cdna clone S3SNUI6s1-36-All 5',
 mRNA sequence.
 BM788452
 BM788452.1 GI:19136684
 EST.
 Homo sapiens (human)
 Homo sapiens
 Mammalia; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 593)
 Kim.N.S., Hahn.Y., Oh.J.H., Lee.J.Y., Ahn.H.Y., Chu.M.Y., Kim.M.R.,
 Oh.K.J., Cheong.J.E., Sohn.H.Y., Kim.J.M., Park.H.S., Kim.S. and
 Kim.Y.S.
 21C Frontier Korean EST Project 2001
 Unpublished
 Contact: Kim YS
 Genome Research Center
 Korea Research Institute of Bioscience & Biotechnology
 52 Boeun-dong Yuseong-gu, Daejeon 305-333, South Korea
 Tel: +82-42-860-4470
 Fax: +82-42-860-4409
 Email: yongsung@mail.kribb.re.kr
 Plate: 36 row: A column: 11
 High quality sequence stop: 593.
 Location/Qualifiers
 1..593
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="S3SNUI6-36-All"
 /sex="F"
 /tissue_type="Ascites"
 /cell_type="Lymphoblast-like"
 /cell_line="SNU-16"
 /lab_host="Top10P"
 /clone_lib="S3SNUI6"
 /note="Organ: Stomach; Vector: pTZ18RP1; Site: 1: EcoRI;
 Site 2: NotI; The poly (A)+ RNA was dephosphorylated with
 bacterial alkaline phosphatase (BAP) and then decapped
 with tobacco acid pyrophosphatase (TAP). The decapped

intact mRNA was ligated with DNA-RNA linker including EcoR
 I site by treatment of T4 RNA ligase and the first strand
 cDNA was synthesized from oligo dT-selected mRNA by
 priming with dT-tailed vector. The dT-tailed vector was
 adjusted to have about 60nt. The cDNA vector was
 circularized with E. coli DNA ligase after digestion of
 EcoRI which site is also included in vector. An RNA strand
 converted to a DNA strand by Okayama-Berg method. The
 obtained cDNA vectors were used for transformation of
 competent cells E. coli Top10p' by electroporation method.
 The cDNA libraries constructed by this method are
 full-length enriched cDNA library."

BASE COUNT 126 a 183 c 154 g 130 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 4,16e-80 Length: 593
 Score: 797.00 Matches: 151
 Percent Similarity: 75.50% Conservative: 0
 Best Local Similarity: 75.50% Mismatches: 0
 Query Match: 49.81% Indels: 49
 DB: 12 Gaps: 1

US-09-607-745-9 (1-292) x BM788452 (1-593)
 QY 86 TrpValLeuThrAlaAlaHisCysPheArgLysHisThrAspValPheAsnTrpLysVal 105
 Db 2 TGGGTCTCTCAGGCACAGCCCATCTTCAGAAACATACCGATGTGTCACTGGAAGGTG 61
 QY 106 ArgAlaGlySerAspLysLeuGlySerPheProSerLeuAlaValAlaLysIleIle 125
 Db 62 CGGGCAGGCTCAGACAAACTGGGCGAGTCCCATCTCTGGCTGTGGCCAGATCATCATC 121
 QY 126 IleGluPheAsnProMetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPhe 145
 Db 122 ATTGAATCAACCCCATGTACCCCAAGACATGATGCGCCCTCATGAGCTGCGATTTC 181
 QY 146 ProLeuThrPheSerGlyThrValArgProIleCysLeuProPhePheAspGluGluLeu 165
 Db 182 CCACCTCACTTTCTCAGGCACAGTCAGGCCCATCTCTCTGCCCTTCTTTGATGAGGAGCTC 241
 QY 166 ThrProAlaThrProLeuThrIleIleGlyTyrGlyPheThrLysGlnAsnGlyGlyLys 185
 Db 242 ACTCAGGCACCCCACTCTGGATCATTTGGATGGGCTTTACGAGCAGAATGGA----- 295
 QY 186 MetSerAspIleLeuLeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAla 205
 Db 295 ----- 295

QY 206 AspAspAlaTyrGlnGlyGluValThrGluLysMetMetCysAlaGlyIleProGluGly 225
 Db 295 ----- 295
 QY 226 GlyValAspThrCysGlnGly-AspSerGlyGlyProLeuMetTyrGlnSerAspGlnTr 245
 Db 296 -----GGGTGACAGTGTGGGGCCCTTGATGATCAATCTGACCAATCTGACCAAGT 337
 QY 245 pHisValValGlyIleValSerTyrGlyTyrGlyCysGlyGlyProSerThrProGlyVa 265
 Db 338 GCATGTGTGGGCATCGTTAGTTGGGCTATGCTGGGGGGCCCGAGCACCCAGGAGT 397
 QY 265 lTyrThrLysValSerAlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu 284
 Db 398 ATACACCAAGGTCTCAGCCTATCTCACTGGATGATCAATGATGCTGGAAGGCTGAGCTG 455

RESULT 14
 M772743
 LOCUS
 DEFINITION
 K-EST0056963 S3SNUI6s1 Homo sapiens cdna clone S3SNUI6s1-22-C01 5',
 mRNA sequence.
 BM772743
 BM772743.1 GI:19102358
 EST.

BM772743 596 bp mRNA linear EST 04-MAR-2002
 K-EST0056963 S3SNUI6s1 Homo sapiens cdna clone S3SNUI6s1-22-C01 5',
 mRNA sequence.
 BM772743
 BM772743.1 GI:19102358
 EST.

BM772743 596 bp mRNA linear EST 04-MAR-2002
 K-EST0056963 S3SNUI6s1 Homo sapiens cdna clone S3SNUI6s1-22-C01 5',
 mRNA sequence.
 BM772743
 BM772743.1 GI:19102358
 EST.

BM772743 596 bp mRNA linear EST 04-MAR-2002
 K-EST0056963 S3SNUI6s1 Homo sapiens cdna clone S3SNUI6s1-22-C01 5',
 mRNA sequence.
 BM772743
 BM772743.1 GI:19102358
 EST.

BM772743 596 bp mRNA linear EST 04-MAR-2002
 K-EST0056963 S3SNUI6s1 Homo sapiens cdna clone S3SNUI6s1-22-C01 5',
 mRNA sequence.
 BM772743
 BM772743.1 GI:19102358
 EST.

BM772743 596 bp mRNA linear EST 04-MAR-2002
 K-EST0056963 S3SNUI6s1 Homo sapiens cdna clone S3SNUI6s1-22-C01 5',
 mRNA sequence.
 BM772743
 BM772743.1 GI:19102358
 EST.

SOURCE
ORGANISM Homo sapiens (human)
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 596)
AUTHORS Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
 Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and
 Kim,Y.S.
TITLE 21C Frontier Korean EST Project 2001
JOURNAL Unpublished
COMMENT Contact: Kim YS
 Genome Research Center
 Korea Research Institute of Bioscience & Biotechnology
 52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea
 Tel: +82-42-860-4470
 Fax: +82-42-860-4409
 Email: yongsung@mail.kribb.re.kr
 Plate: 22 row: C column: 01
 High quality sequence stop: 596.
FEATURES
 source
 1..596
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="S3SNU16s1-22-C01"
 /sex="F"
 /tissue_type="Ascites"
 /cell_type="Lymphoblast-like"
 /lab_host="Top10F"
 /clone_lib="S3SNU16s1"
 /note="Organ: Stomach; Vector: pTZ19RP1; Site: 1: EcoRI;
 Site: 2: NotI; The poly (A)+ RNA was dephosphorylated with
 bacterial alkaline phosphatase (BAP) and then decapped
 with tobacco acid pyrophosphatase (TAP). The decapped
 intact mRNA was ligated with DNA-RNA linker including EcoR
 I site by treatment of T4 RNA ligase and the first strand
 cDNA was synthesized from oligo dt-selected mRNA by
 priming with dr-tailed vector. The dr-tailed vector was
 adjusted to have about 60nt. The cDNA vector was
 circularized with E. coli DNA ligase after digestion of
 EcoRI which site is also included in vector. An RNA strand
 converted to a DNA strand by Okayama-Berg method. The
 obtained cDNA vectors were used for transformation of
 competent cells E. coli Top10F, by electroporation method.
 The cDNA libraries constructed by this method are
 full-length enriched cDNA library. After analyzing and
 sequencing about 2,000 ~ 3,000 colonies in original cDNA
 library, the abundant cDNAs were selected and amplified by
 PCR reaction using vector region primer including T7
 promoter as 5' primer and N(dT)14 as 3' primer. The PCR
 products were used as template for synthesis of
 biotinylated single stranded RNA by in vitro transcription
 reaction. The synthesized RNA probes were hybridized with
 antisense single stranded cDNAs prepared from original
 library and incubated with avidin-gel. After removing
 DNA-RNA hybrids by centrifuge, the subtracted cDNA
 libraries were constructed by transformation of the
 remaining DNA into competent cells E. coli Top10F, with
 electroporation method."

86 TrpValLeuThrAlaAlaHisCysPheArgLysHisThrAspValPheAsnTrpLysVal 105
 2 TGGGTCTCTCAGCGCAGCCACTGCTTCAGGAAACATACCATGTGTTCACCTGAAGGTG 61
 106 ArgAlaGlySerAspLysLeuGlySerPheProSerLeuAlaValAlaLysIleIle 125
 62 CGGGCAGGCTCAGACAACTGGGCGAGCTTCCCATCCCTGGCTGTGGCCAGATCATCATC 121
 126 IleGluPheAsnProMetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPhe 145
 122 ATGTGATTCAACCCCATGTACCCCAAGACATGACATGCCCTCATGAGTGCAGTTC 181
 146 ProLeuThrPheSerGlyThrValArgProLysLeuPhePheAspGluGluLeu 165
 182 CCACCTCCTCTCTCAGGCACAGTCAGGCCCATCTGTCTGCCCTCTCTTCATCAGGAGTTC 241
 166 ThrProAlaThrProLeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyGlyLys 185
 242 ACTCCAGCCACCCCACTCTGGATCATTTGGATGGGCTTTACCAAGCAGATGCA----- 295
 186 MetSerAspIleLeuLeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAla 205
 295 ----- 295
 206 AspAspAlaTyrGlnGlyGluValThrGluLysMetMetCysAlaGlyIleProGluGly 225
 295 ----- 295
 226 GlyValAspThrCysGlnGly-AspSerGlyGlyProLeuMetTyrGlnSerAspGlnTr 245
 296 -----GGGTGACAGTGGTGGGCCCCCTGATGATCAATCTGCACCATCTGACCATGTG 337
 245 ThisValValGlyIleValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyVa 265
 338 GCATGTGGTGGGCATCGTGTAGTGGGGCTATGGCTGGGGGGCCCGAGCACCCAGGAGT 397
 265 TrpThrLysValSerAlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu 284
 398 ATACACCAAGGTCTCAGCCTATCTCAATGGATCTACAAATGCTCGAAGGCTGAGCTG 455
 BM788163 614 bp mRNA linear EST 05-MAR-2002
 K-EST0067267 S3SNU16 Homo sapiens cDNA clone S3SNU16-35-E10 5',
 mRNA sequence.
 BM788163
 VERSION BM788163.1 GI:19136395
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 614)
 AUTHORS Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
 Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and
 Kim,Y.S.
 TITLE 21C Frontier Korean EST Project 2001
 JOURNAL Unpublished
 COMMENT Contact: Kim YS
 Genome Research Center
 Korea Research Institute of Bioscience & Biotechnology
 52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea
 Tel: +82-42-860-4470
 Fax: +82-42-860-4409
 Email: yongsung@mail.kribb.re.kr
 Plate: 35 row: E column: 10
 High quality sequence stop: 614.
 Location/Qualifiers
 1..614
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="S3SNU16-35-E10"

BASE COUNT 127 a 184 c 155 g 130 t
ALIGNMENT
SCORES:
 Score: 4.19e-80 Length: 596
 Percent Similarity: 797.00 Matches: 151
 Conservative: 75.50%
 Local Similarity: 75.50% Mismatches: 0
 Query Match: 49.81% Indels: 49
 Gaps: 12

53-09-607-745-9 (1-292) x BM772743 (1-596)

SE COUNT
IGIN

132 a 189 c 159 g 134 t

/sex="F"
/tissue_type="Ascites"
/cell_type="Lymphoblast-like"
/cell_line="SNU-16"
/lab_host="Top10F"
/clone_lib="S3SNUL6"
/note="Organ: Stomach; Vector: pTZ19RP1; Site 1: EcoRI;
Site 2: NotI; The poly (A)+ RNA was dephosphorylated with
bacterial alkaline phosphatase (BAP) and then decapped
with tabacco acid pyrophosphatase (TAP). The decapped
inact mRNA was ligated with DNA-RNA linker including EcoR
I site by treatment of T4 RNA ligase and the first strand
cDNA was synthesized from oligo dt-selected mRNA by
priming with dt-tailed vector. The dt-tailed vector was
adjusted to have about 60nt. The cDNA vector was
circularized with E. coli DNA ligase after digestion of
EcoRI which site is also included in vector. An RNA strand
converted to a DNA strand by Okayama-Berg method. The
obtained CDA vectors were used for transformation of
competent cells E. coli Top10F' by electroporation method.
The cDNA libraries constructed by this method are
full-length enriched cDNA library."

```

ignment Scores:
ed. No.: 4.39e-80 Length: 614
ore: 797.00 Matches: 151
cent Similarity: 75.50% Conservative: 0
st Local Similarity: 75.50% Mismatches: 0
ery Match: 49.81% Indels: 49
: 12 Gaps: 1
-09-607-745-9 (1-292) x BW789163 (1-614)
86 TrpValLeuThrAlaAlaHisCysPheArgLysHisThrAspValPheAsnTrpLysVal 105
2 TGGGTCTCAGCGCAGCCCATCTCTTCAGGAAACATACCGATGTTTCACTGCGAAGGTG 61
106 ArgAlaGlySerAspLysLeuGlySerPheProSerLeuAlaValAlaLysIleIle 125
62 CGGCGAGGCTCAGACAACTGGGCAGCTTCCCATCCCTGGCTGGCAAGATCATCATC 121
126 IleGluPheAsnProMetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPhe 145
122 ATTGAATTCACCCCATGTCACCCCAAGAACAATGACATCGCCCTCATGAAGTCGCAAGTTC 181
146 ProLeuThrPheSerGlyThrValArgProIleCysLeuProPhePheAspGluGluLeu 165
182 CCACCTCACTTTCTCAGGCACAGTCAGGCCCATCTGCTGCCCTTCTTGGATGAGGAGCTC 241
166 ThrProAlaThrProLeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyGlyLys 185
242 ACTCCAGCCACCCCACTCTGGATCATTGGATGGGGCTTACGGAAGCAGAATGGA----- 295
186 MetSerAspIleLeuLeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAla 205
295 ----- 295
206 AspAspAlaTyrGlnGlyGluValThrGluLysMetMetCysAlaGlyIleProGluGly 225
295 ----- 295
226 GlyValAspThrCysGlnGly-AspSerGlyGlyProLeuMetTyrGlnSerAspGlnTr 245
296 -----GGGTGACAGTGGTGGGGCCCTCGATGTACCAATCTGCACCAAGT 337
245 pHISValValGlyIleValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyVa 265
338 GCATGTGGTGGGCATCGTTAGTTGGGGCTATGGCTGGGGGGCCCCAGACACCCAGAGT 397
265 lTyrThrLysValSerAlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu 284

```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

M protein - nucleic search, using frame_plus_p2n model

run on: November 23, 2003, 07:49:55 ; Search time 391.919 Seconds
(without alignments)
2996.170 Million cell updates/sec

title: US-09-607-745-2

effect score: 2342

sequence: 1 MDPDSQPLNSLDVKPLRPK.....VYTKVSAYLWVWVWKAEL 435

scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

searched: 2552756 seqs, 1349719017 residues

total number of hits satisfying chosen parameters: 5105512

minimum DB seq length: 0

maximum DB seq length: 2000000000

post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

command line parameters:

MODEL=frame+ p2n.model -DEV=xlh
O=/cgn2 1/USPTO.spool/US09607745/runat 21112003 144344 22176/app query.fasta_1.1038
DB=N Geneseq 19Jun03 -QWTF=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPCI=0
LOOPEXT=0 -UNITS=bits -START=1 -END=1 -WATX-blosum62 -TRANS-human40.cdi
LIST=45 -DOCALIGN=200 -THR_SCORE=pct THR_MAX=100 -THR_MIN=0 -ALIGN=15
MODE=LOCAL -OUTFMT=prc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
USER=US09607745.OCGN 1 401 runat 21112003 144344 22176 -NCPU=6 -ICPU=3
NO MAP -LARGQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
GAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N Geneseq 19Jun03.*

1: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*
25: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed.

and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	2342	100.0	1305	22	AAD02556	Human seripancrin
2	2338	99.8	2038	20	AA87154	Human seripancrin HUP
3	2337	99.8	1314	25	ABX76354	Lung cancer-associ
4	2337	99.8	2081	24	AAI72976	CuA8 preferred cDN
5	2337	99.8	2307	24	ABX76529	cDNA encoding huma
6	2335	99.7	2121	24	AAI64284	Human serine prote
7	2329	99.4	1479	22	AAI02557	Human seripancrin
8	2329	99.4	2165	24	ABZ35523	Human gene express
9	2324	99.2	2070	21	AAZ90471	Cancer specific ge
10	2324	99.2	2079	22	AAI13169	Human transmembran
11	2324	99.2	2079	24	AAI72975	cDNA cDNA. Homo s
12	2324	99.2	2079	25	ABX77663	cDNA encoding huma
13	2319	99.0	2137	22	AAI13114	Human membrane-typ
14	2297.5	98.1	2063	21	AAA37099	Human PRO1570 (UNQ
15	2297.5	98.1	2063	22	AA546089	Human DNA encoding
16	2297.5	98.1	2063	22	AAF92113	Human PRO1570 cDNA
17	2297.5	98.1	2063	22	AA543396	DNA encoding prote
18	2297.5	98.1	2063	24	ABX74433	Human cDNA encodin
19	2297.5	98.1	2063	24	ABK11090	cDNA encoding tumo
20	2297.5	98.1	2063	25	ACA57847	Human PRO1570 cDNA
21	2297.5	98.1	2063	25	ACA58865	cDNA encoding huma
22	2297.5	98.1	2063	25	ACA60418	Novel human secret
23	2297.5	98.1	2063	25	ACA63428	cDNA encoding huma
24	2297.5	98.1	2063	25	ABX98317	Human cDNA encodin
25	2297.5	98.1	2063	25	ABX98819	Novel human secret
26	2297.5	98.1	2063	25	ACA05864	Human secreted/tra
27	2297.5	98.1	2063	25	ABX97908	Human PRO polynucl
28	2297.5	98.1	2063	25	ABX78692	Human PRO polynucl
29	2297.5	98.1	2063	25	ABX75705	Human cDNA encodin
30	2297.5	98.1	2063	25	ABX76910	Human PRO polynucl
31	2297.5	98.1	2063	25	ABX16750	Human cDNA encodin
32	2251	96.1	1795	22	AA546880	Human cDNA encodin
33	1668	76.1	1854	22	AAH99574	Human protein enco
34	1287	55.0	699	18	AAI79127	Human serine prote
35	1286	54.9	1281	21	AAA97361	Human colorectal c
36	1259	53.8	1189	24	AAI64290	Protease D-G catay
37	1091	46.6	649	22	AAH33477	Human colon cancer
38	1090	46.5	867	22	AAH30210	DNA encoding rena
39	1090	46.5	867	22	AAH26952	Human cDNA encodin
40	1090	46.5	867	22	AAH33181	DNA encoding huma
41	918	39.2	1084	22	AAH34953	Human colon cancer
42	806	34.4	1008	20	AAH04381	Human secreted pro
43	699.5	29.9	2067	22	AAH05797	Human transmembran
44	698.5	29.8	1434	25	ABZ22864	Human dendritic ce
45	697.5	29.8	1314	22	AAH83971	Human SER6 nucleot

ALIGNMENTS

RESULT 1
AAD02556
ID AAD02556 standard; cDNA; 1305 BP.
XX
AC AAD02556;
XX
DT 02-MAY-2001 (first entry)
XX
DE Human seripancrin cDNA.

Human; seripancrin; serine protease; chromosome 11q22-q23; therapy;
arthritis; chronic obstructive pulmonary disorder; COPD; cancer;
osteoporosis; aberrant wound healing; angiogenesis; diabetes;
inflammatory disorder; stroke; cardiovascular disease; gene therapy;
vaccine; cytostatic; cerebroprotective; vulnerable; osteopathic; ss.
Homo sapiens. OS

1261 TCAGCCTATCTCACTGGAATCTACAAATGCTGGAGGCTGAGCTG 1305

RESULT 2
D AX87154
D AX87154 standard; cDNA; 2038 BP.
C AX87154;
X 27-SEP-1999 (first entry)
T Human protease HUPM-6 cDNA.
X Serine protease; human; HUPM-6; cell proliferation; cancer;
X immune disorder; inflammation; therapy; ss.
X Homo sapiens.
X Key Location/Qualifiers
X CDS. 200..1507
X /*tag= a
X W09936550-A2.
X 22-JUL-1999.
X 12-JAN-1999; 99WO-US00655.
X 16-JAN-1998; 98US-0008271.
X (INCY-) INCYTE PHARM INC.
X Bandman O, Corley NC, Guegler KJ, Hillman JL, Shah P;
X Tang YT, Yue H;
X WPI; 1999-430616/36.
X P-PSDB; AAY06437.
X Novel human protease molecules useful in the treatment of
X developmental disorders and/or cancers
X Claim 8; Page 86-87; 90pp; English.
X This nucleotide sequence codes for HUPM-6 (see AAY06437), a novel
X human protease. HUPM-6 cDNA was initially identified in Incyte
X Clone 1337018 from the colon cDNA library COLMN013 using a
X computer search for amino acid sequence alignments. The present
X sequence is a consensus sequence derived from overlapping and/or
X extended nucleic acid sequences: Incyte Clones 1271725 (TGSTTUT02),
X 1337018, 586982 and 589598 (UTRSN0701). A fragment comprising
X nucleotides 900-949 of the present sequence can be used for
X hybridisation. This sequence encompasses an active site residue.
X Northern analysis shows expression of HUPM-6 in gastrointestinal,
X and male an female reproductive cDNA libraries. Approximately 65%
X of these libraries are associated with neoplastic disorders, and
X 22% with the immune response. The invention provides 12 new human
X proteases, i.e. HUPM-1 to -12 (see AAY06432-43), and the
X polynucleotides encoding them (see AAX87149-60). Also provided are
X vectors, host cells and methods for producing HUPM polypeptides, as
X well as agonists and antagonists of HUPM. Methods for treating or
X preventing cell proliferative disorders and immune disorders using
X HUPM or HUPM antagonists are claimed.
X Sequence 2038 BP; 462 A; 591 C; 569 G; 416 T; 0 other;
X
X Alignment Scores:
X Ed. No.: 2.72e-209 Length: 2038
X ore: 2338.00 Matches: 434
X cent Similarity: 100.00% Conservative: 1
X st Local Similarity: 99.77% Mismatches: 0
X ery Match: 99.83% Indels: 0
X Gaps: 20

US-09-607-745-2 (1-435) x AAX87154 (1-2038)

Qy 1 MetAspProAspSerAspGlnProLeuAsnSerLeuAspValLysProLeuArgLysPro 20
Db 200 ATGGATCTCTGACAGTATCAACCTCTGAAACAGCCTGATGCAACCCCTGGCAACCC 259
Qy 21 ArgIleProMetGluThrPheArgLysValGlyLeuProIleIleIleAlaLeuSer 40
Db 260 CGTATCCCATGGAGACTTCAGAAAGTGGGATCCCATCATCATAGCAGTACTGAGC 319
Qy 41 LeuAlaSerIleIleIleValValValLeuIleLysValIleLeuAspLysTyrPhe 60
Db 320 CTGGCAGTATCATATTGTGGTTGCTCATCAAGGTGATTCGTGATAAATACTACTTC 379
Qy 61 LeuCysGlyGlnProLeuHisPheIleProArgLysGlnLeuCysAspGlyGluLeuAsp 80
Db 380 CTCTCGGGGAGCCTCTCCACTTCATCCGAGGAAGCAGCTGTGTGACGAGAGCTGGAC 439
Qy 81 CysProLeuGlyGluAspGluHisCysValLysSerPheProGluGlyProAlaVal 100
Db 440 TGTCTCTGGGGAGGAGGAGGAGCACTGTGTCAAGAGCTTCCCGAGAGGCTTGCAGTG 499
Qy 101 AlaValArgLeuSerLysAspArgSerThrLeuGlnValLeuAspSerAlaThrGlyAsn 120
Db 500 GCAGTCCGCTCTCCAAAGGACCGATCCACACTGCAGGTGTGGACTCGGCCACAGGGAAC 559
Qy 121 TrpPheSerAlaCysPheAspAsnPheThrGluAlaLeuAlaGluThrAlaCysArgGln 140
Db 560 TGGTCTCTGCTGCTTTCGACAACTTCACAGAAGCTCTCGCTGAGACAGCTGTAGCAG 619
Qy 141 MetGlyTyrSerSerLysProThrPheArgAlaValGluIleGlyProAspGlnAspLeu 160
Db 620 ATGGGCTACAGCAGCAAAACCCACTTTTCAGAGCTGTGGAGATTGGCCACAGACCAAGATCTG 679
Qy 161 AspValValGluIleThrGluAsnSerGluGluLeuArgMetArgAsnSerSerGlyPro 180
Db 680 GATGTTGTTGAATCACAGAAACAGCCAGGAGCTTCGCATGCGGAACTCAAGTGGGCC 739
Qy 181 CysLeuSerGlySerLeuValSerLeuHisCysLeuAlaCysGlyLysSerLeuLysThr 200
Db 740 TGTCTCTCAGGCTCCCTGGTCTCCCTGCACCTGTCTTGCCTGTGGGGAGAGCTGAAGACC 799
Qy 201 ProArgValValGlyGlyGluGluAlaSerValAspSerTrpProTrpGlnValSerIle 220
Db 800 CCCGCTGTGTGGTGGGGAGGAGGCTCTGTGGATTCTTGGCCTTGGCAGGTGAGATC 859
Qy 221 GlnTrpAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThr 240
Db 860 CAGTACGACAAACAGCAGCTGTGTGGGGAGGAGCACTCTGGACCCCTGGTCTCTCAG 919
Qy 241 AlaAlaHisCysPheArgLysHisThrAspValPheAsnTrpLysVala:GAlaGlySer 260
Db 920 GCAGCCCACTGCTTCAGAAACATACATACGATGTGTTCACCTGGAAGGTGGCGGAGGTCA 979
Qy 261 AspLysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleGluPheAsn 280
Db 980 GACAAACTGGGCGAGCTTCCCATCCCTGGCTGTGGCCCAAGATCATCATGAAATTCAC 1039
Qy 281 ProMetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPhe 300
Db 1040 CCCATGTACCCCAAGACAATGACATCGCCCTCATGAAGCTGAGTTCACACTCCTTTC 1099
Qy 301 SerGlyThrValArgProIleCysLeuProPhePheAspGluLeuLeuThrProAlaThr 320
Db 1100 TCAGGCACAGTCAGGCCCATCTGTCTGCCCTTCTTGTATGAGGAGCTCACTCCAGCCACC 1159
Qy 321 ProLeuTyrPheIleGlyTrpGlyPheThrLysGlnAsnGlyGlyLysMetSerAspIle 340
Db 1160 CCACCTCTGATCATTGATGGGGCTTTACGAGAGCAAGATGGAGGAAGATGCTGACATA 1219
Qy 341 LeuLeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAspAlaTyr 360
Db 1220 CTGCTGCAGGGCTCAGTCCAGGTCATTGACAGCACCGGTGCAATGCAGACGATGCGTAC 1279

Y 361 GlnGlyGluValThrGluLysMetMetCysAlaGlyLeProGluGlyGlyValAspThr 380
b 1280 CAGGGGGAAGTCCCGGAGAGATGATGTGCGAGCATCCCGAAGGGGTGTGGACAC 1339
Y 381 CysGlnGlyAspSerGlyGlyProLeuMetTyrglnSerAspGlnTrpHisValValGly 400
b 1340 TGCAGGGGTGACAGTGGTGGCCCTGATGTACCAATCTGACCAAGTGGCAATGGTGGGC 1399
Y 401 IleValSerTrpGlyTyrglyCysGlyGlyProSerThrProGlyValTyThrLysVal 420
b 1400 ATCGTTAGCTGGGCTATGCTGGGGGGCCGAGCACCCAGGAGTATACACCAAGTTC 1459
Y 421 SerAlaTyrlleuAsnTrpIleTyrglnValTrpLysAlaGluLeu 435
b 1460 TCAGCCTATCTCACTGGATCTACAATGTCTGGAAGGCTGAGCTG 1504

RESULT 3

BX76354

D ABX76354 standard; DNA; 1314 BP.

X X X

ABX76354;

02-APR-2003 (first entry)

Lung cancer-associated polynucleotide #218.

Lung cancer-associated polynucleotide; gene; ds; cytostatic; emphysema;
antiinflammatory; antiasthmatic; non-small cell lung cancer; atelectasis;
small cell lung cancer; benign lesion; precancerous lesion; bronchitis;
chronic obstructive pulmonary disease; hypersensitivity pneumonitis;
interstitial pulmonary fibrosis; fibrosis; asthma; bronchiectasis.

Unidentified.

WO200286443-A2.

31-OCT-2002.

18-APR-2002; 2002WO-US12476.

18-APR-2001; 2001US-284770P.

10-MAY-2001; 2001US-290492P.

09-NOV-2001; 2001US-339245P.

13-NOV-2001; 2001US-350666P.

29-NOV-2001; 2001US-334370P.

12-APR-2002; 2002US-372246P.

(BOSB-) EOS BIOTECHNOLOGY INC.

Aziz N, Murray R;

WPI; 2003-093161/08.

P-PSDB; ABU56625.

Detecting a lung cancer-associated transcript in a cell from a patient
for treating lung cancer, by contacting a biological sample from the
patient with a polynucleotide that exhibits increased or decreased
expression in lung cancer -

Claim 22; Page 353; 453pp; English.

The invention relates to a method for detecting a lung cancer-associated
transcript in a cell from a patient, comprising contacting a biological
sample from the patient with a polynucleotide that selectively hybridizes
to a sequence that is at least 80 % identical to a gene that exhibits
increased or decreased expression in lung cancer samples. Lung
cancer-associated polynucleotides and polypeptides are used for
identifying a compound that modulates a lung cancer-associated
polypeptide, for inhibiting proliferation of a lung cancer-associated
cell to treat lung cancer in a patient and for treating a mammal having
lung cancer by administering a modulatory compound identified. The
methods are useful for treating lung cancer, such as small cell lung

CC cancer, non-small cell lung cancer or other benign or precancerous
CC lesions, e.g. atelectasis, emphysema, bronchitis, chronic obstructive
CC pulmonary disease, fibrosis, hypersensitivity pneumonitis, interstitial
CC pulmonary fibrosis, asthma and bronchiectasis. The genes, polynucleotides
CC and polypeptides are useful for diagnostic purposes and as targets for
CC screening for therapeutic compounds that modulate lung cancer, such as
CC antibodies. Sequences ABX76124-ABX76474 represent lung cancer-associated
CC polynucleotides of the invention.

XX

SQ Sequence 1314 BP; 297 A; 369 C; 373 G; 275 T; 0 other;

Alignment Scores:

Pred. No.:	1.88e-209	Length:	1314
Score:	2337.00	Matches:	434
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	99.79%	Indels:	0
DB:	25	Gaps:	0

US-09-607-745-2 (1-435) x ABX76354 (1-1314)

QY 2 AspProAspSerAspGlnProLeuAsnSerLeuAspVallyProLeuArgLysProArg 21

Db 10 GATCTGACAGTGAATCACTCTGACAGCTTCGATGTAACCCCTGGCAACCCCGT 69

QY 22 IleProMetGluThrPheArgLysValGlyIleProIleIleAlaLeuLeuSerLeu 41

Db 70 ATCCCATGGAGACCTTCAGAAAGGTGGGATCCCATCATCATATGACTACTGAGCCTG 129

QY 42 AlaSerIleIleIleValValValIleValIleValIleValIleAspLysTrpPheLeu 61

Db 130 GCGAGTATCATATTGTGTCTCTCATCAAGGTGATTCGATAAATACTACTTCTCTC 189

QY 62 CysGlyGlnProLeuHisPheIleProArgLysGlnLeuCysAspGlyGluLeuAspCys 81

Db 190 TGGGGGACGCTCTCCACTTCATCCGAGAGACAGCTGTGTGACGAGCTGGAGCTGT 249

QY 82 ProLeuGlyGluAspGluGluHisCysVallySerPheProGluGlyProAlaValAla 101

Db 250 CCGTTGGGGGAGGACGAGGAGCACGTGTCAAGAGCTTCCCCGAGGGGCTGCAGTGGCA 309

QY 102 ValArgLeuSerLysAspArgSerThrLeuGlnValLeuAspSerAlaThrGlyAsnTrp 121

Db 310 GTCCGCTCTCCAGGACCGATCCACTGCGAGTGTGCTGCGCCACAGGGAACCTGG 369

QY 122 PheSerAlaCysPheAspAsnPheThrGluAlaLeuAlaGluThrAlaCysArgGlnMet 141

Db 370 TTCTCTGCTCTTTCGACACCTTCAGAGACTCTGCTGAGACACCTGTAGGCAATG 429

QY 142 GlyTrpSerSerLysProThrPheArgAlaValGluIleGlyProAspGlnAspLeuAsp 161

Db 430 GGGTACAGCAGCAAAACCCACTTTCAGAGCTGTGGAGATTGGCCAGACAGGATCTGGAT 489

QY 162 ValValGluIleThrGluAsnSerGlnGluLeuArgMetArgAsnSerSerGlyProCys 181

Db 490 GTTGTGTAATCACAGAAACAGCCAGGAGCTTCGCATCGGAACTCAAGTGGGCCCTGT 549

QY 182 LeuSerGlySerLeuValSerLeuHisCysLeuAlaCysGlyLysSerLeuLysThrPro 201

Db 550 CTCTCAGGCTCCCTGGTCTCCCTGACACTGTCTGCTGTGGGAAGAGCCTGAAGACCCC 609

QY 202 ArgValValGlyGlyGluGluAlaSerValAspSerTrpProTrpGlnValSerIleGln 221

Db 610 CGTGTGGTGGGTGGGAGGAGGCTCTGTGGATTCTTGGCCCTTGGCAGTCAAGATCCAG 669

QY 222 TyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThrAla 241

Db 670 TACGCAAAACAGCAGTCTGTGGAGGAGCATCTCTGAGACCCCTGAGCTGCTCAGGCA 729

QY 242 AlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySerAsp 261

Db 730 GCCCACTGCTTCAGGAACATACCCGATGTGTTCAACTGGAGGTGGGGCAGGCTCAGAC 789

262 LysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleIleGluPheAsnPro 281
 790 AAATGGGGAGCTTCCCATCCCTGGCTGGCCAGATCATCATTAATCAACCC 849
 282 MetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPheSer 301
 850 ATGTACCCCAAGACATGACATCGCCCTCATGAGCTGAGTTCACCTCACTTCTCA 909
 302 GlyThrValArgProIleCysLeuProPhePheAspGluLeuThrProAlaThrPro 321
 910 GGCACAGTCAAGCCCATCTGCTGCTCCCTTCTTTCATGAGGAGCTCACTCCAGCCACCCCA 969
 322 LeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyCysMetSerAspIleLeu 341
 970 CTCGTGATCATGATGGGCTTTTACGAGCAGATGAGGGAGATGCTGACATAC 1029
 342 LeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAlaThrGln 361
 1030 CTCGAGCGCTCAGTCCAGGTCATTGACAGCACACGGTGCATGCAGACGATGCTAC 1089
 362 GlyGluValThrGlnLysMetMetCysAlaGlyIleProGluGlyValAspThrCys 381
 1090 GGGGAGTCAACGAGAGATGATGTGTGACGATCCCGAAGGGGTGTGGACACCTGC 1149
 382 GlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGlyIle 401
 1150 CAGGTGACAGTGGTGGCCCTGATGTACCAATCTGACAGTGGCATGTGGTGGCATC 1209
 402 ValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyValThrLysValSer 421
 1210 GTTAGCTGGGGCTATGCTCGGGGGCCCGAGCACCCCGAGGATATACCAAGGTCTCA 1269
 422 AlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu 435
 1270 GCCTATCTCACTGGATCTACATGCTGTGGAAGGCTGAGCTG 1311

SULT 4
 172976
 AAI72976 standard; cDNA; 2081 BP.

AAI72976;

21-AUG-2002 (first entry)

CUA8 preferred cDNA.

Gene; colorectal cancer; CGA7; CUA8; modulating protein; screening;
 drug candidate; vaccine; ss.

Homo sapiens.

Key Location/Qualifiers
 CDS 215..1528
 /*tag= a
 /product= "CUA8"

US2002042067-A1.

11-APR-2002.

08-MAY-2001; 2001US-0851588.

17-AUG-2000; 2000US-0642252.

06-SEP-2000; 2000US-0656002.

(MACK/) MACK D.
 (GISH/) GISH K C.
 (WILSON/) WILSON K E.

Mack D, Gish KC, Wilson KE;

WPI; 2002-453647/48.

P-PSDB; AAG79359.

XX Screening drug candidates for treating colorectal cancer, comprises
 PT determining the effect of the candidate on the expression profile gene
 PT of CGA7 or CUA8 -

Claim 69; Fig 7; 40pp; English.

XX The sequences given in AAI72973-76 encode the colorectal cancer
 CC proteins, CGA7 and CUA8. The CGA7 and CUA8 proteins are colorectal
 CC cancer modulating proteins and have been mapped to chromosomes 2 (CGA7)
 CC and 11 (CUA8). These sequences may be used in the method of the
 CC invention for screening drug candidates. The method comprises adding
 CC a drug candidate to a cell that expresses an expression profile gene
 CC encoding CGA7, CUA8 or fragments and determining the effect of the
 CC drug candidate on the expression of the expression profile gene. The new
 CC methods are used to screen bioactive agents for the ability to bind to
 CC or modulate the activity of CGA7 or CUA8 and evaluate the effect of a
 CC candidate colorectal cancer drug. An antibody to CGA7 or CUA8 can inhibit
 CC the activity of CGA7 or CUA8, respectively, and is used to screen for
 CC an agent that can interfere with the binding of CGA7 or CUA8 to the
 CC antibody. The antibody can be used to treat colorectal cancer. The
 CC antibody or a fragment of it is used to localize a therapeutic group to
 CC a colorectal cancer tissue, where the therapeutic group is a cytotoxic
 CC agent or a radioisotope. Antisense molecules are used to inhibit
 CC colorectal cancer in a cell. Nucleic acid segments encoding CGA7 or
 CC CUA8 are used in a biochip. CGA7, CUA8 or a nucleic acid encoding it
 CC are used to elicit an immune response. CGA7 or CUA8 is used to
 CC determine the prognosis of an individual with colorectal cancer.
 CC Nucleic acid encoding CGA7 or CUA8 can be used in vaccines.

XX Sequence 2081 BP; 484 A; 597 C; 576 G; 424 T; 0 other;

Alignment Scores:

Pred. No.:	3,47e-209	Length:	2081
Score:	2337.00	Matches:	434
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	99.79%	Indels:	0
DB:	24	Gaps:	0

US-09-607-745-2 (1-435) X AAI72976 (1-2081)

QY 2 AspProAspSerAspGlnProLeuAsnSerLeuAspValLysProLeuArgLysProArg 21
 Db 224 GATCTGACAGTATCAACCTCTGACAGCTCGATGCAACCCCTGCGCAACCCCGT 283
 QY 22 IleProMetGluThrPheArgLysValGlyIleProIleIleIleAlaLeuSerLeu 41
 Db 284 ATCCCTTGGAGACCTTCAGAAAGTGGGATCCCATCATCATAGCCTACTAGGCTG 343
 QY 42 AlaSerIleIleIleValValValLeuIleLysValIleLeuAspLysTyrTyrPheLeu 61
 Db 344 GCGAGTATCATATTGCTGCTCATCAAGTGATCTCGGATAAATACTACTTCTCTC 403
 QY 62 CysGlyGlnProLeuHisPheIleProArgLysGlnLeuCysaspGlyGluLeuAspCys 81
 Db 404 TCGGGGAGCCCTCTCCACTTCATCCGAGGAAGCAGCTGTGTGACGAGAGCTGGACTGT 463
 QY 82 ProLeuGlyGluAspGluGluHisCysValLysSerPheProGluGlyProAlaValAla 101
 Db 464 CCTTTGGGGAGGACGAGGAGCACTGTGTCAAGAGCTTCCCGAGGGCTTCAGTGGCA 523
 QY 102 ValArgLeuSerLysAspArgSerThrLeuGlnValLeuAspSerAlaThrGlyAsnTrp 121
 Db 524 GTCCGCTCTCCAAAGGACCGATCCACACTGCTGGACTCGGCCACAGGGAACCTGG 583
 QY 122 PheSerAlaCysPheAspAsnPheThrGluAlaLeuAlaGluThrAlaCysArgGlnMet 141
 Db 584 TTCTCTGCTGTGTTTCGACAACTTCACAGAAGCTCTCGCTGAGACAGCCTGTAGCGAGATG 643
 QY 142 GlyTyrSerSerLysProThrPheArgAlaValGluIleGlyProAspGlnAspLeuAsp 161
 Db 644 GGCTACAGCAGCAGCAACCCACTTTCAGAGCTGTGGAGATTGGCCCGAGGATCTGGAT 703

Y 162 ValValGluLeuThrGluAsnSerGlnGluLeuArgMetArgAsnSerSerGlyProCys 181
b 704 GTTGTGAAATTCACAGAAAAACAGCAGAGCTTCGCATGGGAACTCAAGTGGGCCCTGT 763
Y 182 LeuSerGlySerLeuValSerLeuHisCysLeuAlaCysGlyLysSerLeuLysThrPro 201
b 764 CTCTCAGGCTCCCTGGTCTCCCTGCATCTCTTCCCTGTGGGAAAGCCCTGAAGACCCCC 823
Y 202 ArgValValGlyGlyGluGluAlaSerValAspSerTrpProTrpGlnValSerIleGln 221
b 824 CGTGTGTGTGGTGGGAGGAGGCTCTGTGATCTTGGCTTGGCAGGTCAGCATCCAG 883
Y 222 TyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisIleTrpValLeuThrAla 241
b 884 TAGACAAAACAGCAGCTGTGTGGAGGAGCATCTGGACCCCACTGGGTCTCACGGCA 943
Y 242 AlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySerAsp 261
b 944 GCCACTGCTTCAGGAACATACCATGCTGTTCACCTGGAGAGTGGCGGACGGCTCAGAC 1003
Y 262 LysLeuGlySerPheProSerLeuAlaValAlaLysIleIleLeuGluPheAsnPro 281
b 1004 AAACCTGGCAGCTTCCCATCCCTGGCTGTGGCCAAAGATCATCATTAATTCAACCCC 1063
Y 282 MetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPheSer 301
b 1064 ATGTACCCCAAGAACATGATGATCCCTCATGAGCTGAGTTCACACTCTTCTCA 1123
Y 302 GlyThrValArgProLysCysLeuProPheAspGluGluLeuThrProAlaThrPro 321
b 1124 GGCACAGTCAGGCCCATCTCTCTGCCCTCTTTGATGAGGAGCTCACTCCAGCCACCCA 1183
Y 322 LeuTrpIleLeuGlyTrpGlyPheThrLysGlnAsnGlyLysMetSerAspIleLeu 341
b 1184 CTCTGGATCATTTGATGGGCTTTACGAAGCAGAAATGGAGGAGATGTCTGCACATCTG 1243
Y 342 LeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAspAlaTyrGln 361
b 1244 CTGAGCGCTCAGTCCAGGTCATTTGACAGACACGGTGCATGCAGACGATGCTACAC 1303
Y 362 GlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyLysValAspThrCys 381
b 1304 GGGGAAGTCAACGAGAGATGATGTGTGAGGATCCCGAGAGGGGTGTGACACTGC 1363
Y 382 GlnGlyAspSerGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGlyIle 401
b 1364 CAGGGTGACAGTGGTGGCCCTCTGATGTACCAATCTGACCACTGGCATGTGTGGCATC 1423
Y 402 ValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyValTyrThrLysValSer 421
b 1424 GTTAGCTGGGGCTAATGCTCGGGGGCCGAGACCCCGAGAGTATACCAAGGTCTCA 1483
Y 422 AlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu 435
b 1484 GCCTACTCAACTGGATCTACATGCTCGAAGGCTGAGCTG 1525

RESULT 5

D BS76529
X ABS76529 standard; cDNA; 2307 BP.

C ABS76529;

T 11-DEC-2002 (first entry)

X cDNA encoding human ovarian cancer marker OV86.

W Human; ovarian cancer; marker; cancer; familial history; brain disorder;
W central nervous system disorder; bacterial meningitis; viral meningitis;
W Alzheimer's disease; Parkinson's disease; cerebral oedema; hydrocephalus;
W brain herniation; inflammation; encephalitis; testicular disorder;
W nontuberculous granulomatous orchitis; connective tissue disorder;
W heart disorder; ischaemic heart disease; atherosclerosis; neoplasm;

histological type; carcinogenic; ovarian cancer marker; gene; ss.
Homo sapiens.

WO200271928-A2.

19-SEP-2002.

14-MAR-2002; 2002WO-US07826.

14-MAR-2001; 2001US-276025P.

14-MAR-2001; 2001US-276026P.

10-AUG-2001; 2001US-311732P.

19-SEP-2001; 2001US-323580P.

26-SEP-2001; 2001US-324967P.

26-SEP-2001; 2001US-325102P.

26-SEP-2001; 2001US-325149P.

(MILL-) MILLENNIUM PHARM INC.

Morahan JE, Gannavarapu M, Hoersch S, Kamatkar S, Kovatis SG;

Meyers RE, Morrissey MP, Olandt PJ, Sen A, Vieby PO, Mills GB;

Bast RC, Lu K, Schmandt RE, Zhao X, Glatt K;

WPI; 2002-723277/78.

P-PSDB; AB96430.

Assessing whether a patient is afflicted with ovarian cancer, useful in

assessing the stage or progression of the disease, comprises comparing the expression level of a cancer marker in a sample from a patient and from a non cancer patient -

Disclosure; Page 438; 481pp; English.

The present invention relates to a new method for assessing whether a patient is afflicted with ovarian cancer. The method involves comparing the expression level of a marker in a patient sample and the normal level of expression of the marker in a control non-ovarian cancer sample, where the marker is selected from 363 cancer markers described in the specification. The method of the invention is useful in diagnosing or characterising cancer, in detecting the presence of cancer as early as possible, and the recurrence of ovarian cancer. The method may also be of particular use with patients having an enhanced risk of developing ovarian cancer (e.g. patients having a familial history of ovarian cancer). The cancer markers may be used in the management and treatment of e.g. brain and central nervous system disorders (e.g. bacterial and viral meningitis, Alzheimer's disease or Parkinson's disease), brain disorders (e.g. cerebral oedema, hydrocephalus or brain herniations), inflammations (e.g. bacterial or viral meningitis or encephalitis), testicular disorders (e.g. nontuberculous granulomatous orchitis), connective tissue disorders, or heart disorders (e.g. ischaemic heart disease or atherosclerosis). The compositions and methods may also be used in assessing the histological type of neoplasm associated with ovarian cancer, monitoring the progression of ovarian cancer, determining whether ovarian cancer has metastasized or is likely to metastasize, selecting a composition for inhibiting ovarian cancer, assessing the ovarian carcinogenic potential of a compound, or inhibiting ovarian cancer or at risk of developing ovarian cancer. The present nucleic acid sequence encodes one of the ovarian cancer markers described in the invention.

SQ Sequence 2307 BP; 555 A; 647 C; 614 G; 460 T; 31 other;

Alignment Scores:

Pred. No.:	3-986-209	Length:	2307
Score:	2337.00	Matches:	434
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	99.79%	Indels:	0
DB:	24	Gaps:	0

US-09-607-745-2 (1-435) x AB576529 (1-2307)

2 AspProAspSerSerGlnProLeuAsnSerLeuAspVallySProLeuArglySPDRg 21
 284 GATCTCAGAGTATCAACCTCTGAACAGAGCTCGATGTAACACCCCTGGCAACCCCGT 343
 22 IleProMetGluThrPheArgLysValGlyIleProIleIleLeuLeuSerLeu 41
 344 ATCCCATGGAGACCTTCAGAAAGGTGGGATCCCATCATCATAGACTACTGAGCCG 403
 42 AlaSerIleIleLeuValValLeuIleLysValIleLeuAspLysPheThrPheLeu 61
 404 GCGAGTATCATATTGTTGCTTCATCAAGGTGATTCTGGATAAATACTACTTCCTC 463
 62 CysGlyGlnProLeuHisPheIleProArgLysGlnLeuCysAspGlyGlnLeuAspCys 81
 464 TGGGGGAGGAGCTTCACCTTCATCCCGAGAGAGCTGTGTACGAGAGCTGAGCT 523
 82 ProLeuGlyGluAspGluGluHisCysValLysSerPheProGluGlyProAlaValAla 101
 524 CCCTTGGGGGAGGAGGAGGAGCTGTGTCAAGAGCTTCCCGAAGGCTTCAGTGGCA 583
 102 ValArgLeuSerLysAspArgSerThrLeuGlnValLeuAspSerAlaThrGlyAsnTrp 121
 584 GTCCGCTCTCCAAAGGACCGATCCACACTCGAGGTGCTGAGCTCGGCCACAGGGAAC 643
 122 PheSerAlaCysPheAspAsnThrGluAlaLeuAlaGluThrAlaCysArgGlnMet 141
 544 TTCTCTGCTGTTTCGAACTTCACAGAGCTCTCGCTGAGACACCTGTAGGACAGATG 703
 142 GlyThrSerLysProThrPheArgAlaValGluIleGlyProAspGlnAspLeuAsp 161
 704 GGCATACAGACAGAACCCACTTTTCAGAGCTGTGAGATTGGCCGACAGCAGATCTGAT 763
 162 ValValGluIleThrGluAsnSerGlnGluLeuArgMetArgAsnSerSerGlyProCys 181
 764 GTTGTGAAATACAGAAACAGCCAGAGAGCTTCGATGCGGAACTCAAGTGGGCCCTGT 823
 182 LeuSerGlySerLeuValSerLeuHisCysLeuAlaCysGlyLysSerLeuLysThrPro 201
 824 CTCTCAGGCTCCCTGGTCTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 883
 202 ArgValValGlyGlyGluGluAlaSerValAspSerTrpProGlnValSerIleGln 221
 884 CGTGTGTGGTGGGAGAGGCTCTGTGTGATTTCTTGGCTTGGCAGGTGACATCCAG 943
 222 TyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThrAla 241
 944 TACGACAAACAGCAGCTCTGTGAGGAGGAGCATCTGGACCCCACTGGGTCTCAGGGCA 1003
 242 AlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySerAsp 261
 1004 GCCCACTCTTCAGAAACATACCGATGTGTTCACCTGGAGGTGGGCGAGCTCAGAC 1063
 262 LysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleLeuGluPheAsnPro 281
 1064 AAATGGGAGCTTCCATCCCTGCTGTGGCCAAAGATCATCATCATTTGAATTCACCC 1123
 282 MetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPheSer 301
 1124 ATGTACCCCAAGACAATGACATCGCCCTCATGAAGCTGAGTTCACACTCTCTCTCA 1183
 302 GlyThrValArgProIleCysLeuProPhePheAspGluGluLeuThrProAlaThrPro 321
 1184 GGCAGTCAAGCCCATCTGTCTGCCCTTCTTGTATGAGGAGCTCACTCCAGCCACCCCA 1243
 322 LeuThrPheIleGlyThrGlyPheThrLysGlnAsnGlyLysMetSerAspIleLeu 341
 1244 CTCTGGATCATTTGATGGGCTTTCAGAAAGCAGATGAGGAGGAGATGCTGACATCTG 1303
 342 LeuGlnAlaSerValGlnValIleAsnSerThrArgCysAsnAlaAspAspAlaThrGln 361
 1304 CTGAGGCGTCACTGCTCAGTTCATGACAGACAGCGGTGCAATGACAGAGTGGTACAG 1363
 362 GlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThrCys 381

1364 GGGGAAGTACCCAGAGAGATGATGTGAGGATCCCGAAGGGGTGGACACCTGC 1423
 382 GlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGlyIle 401
 1424 CAGGTGACAGTGTGGGCCCTGATGTACCAATCTGACCACTGGCATGTGTGGGCATC 1483
 402 ValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyValThrLysValSer 421
 1484 GTTAGCTGGGCTATGCTGCGGGGCGCGACCCAGGAGTATACACCAAGGTCTCA 1543
 422 AlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu 435
 1544 GCCTACTCTCACTGATCTACATATGCTCGAAGGCTGAGCTG 1585

RESULT 6
 AAI64284
 ID AAI64284 standard; cDNA; 2121 BP.
 XX AAI64284;
 AC AAI64284;
 XX 07-MAY-2002 (first entry)
 DT Human serine protease D-G cDNA.
 DE Human serine protease D-G cDNA.
 XX Serine protease; D-G; human; zymogen; enzyme; cytostatic;
 KW anti-inflammatory; dermatological; anticoagulation; cancer;
 KW skin disorder; neuropathic pain; inflammatory disorder;
 KW coagulation diathesis; thrombosis; laundry detergent; skin care;
 KW gene therapy; ss.
 XX Homo sapiens.
 OS Homo sapiens.
 FH Key
 CDS 277..1584
 FT /*tag= a
 FT /product= "serine protease D-G"
 FT /transl_except= (pos: 1357..1359, aa: Gln)
 FT polyA_signal 2100..2105
 FT /*tag= b
 FT /note= "does not conform to consensus sequence"
 FT
 XX WO200202011-A1.
 XX 10-JAN-2002.
 XX 08-JUN-2001; 2001WO-US18568.
 XX 30-JUN-2000; 2000US-0607745.
 XX (ORTH) ORTHO-MCNEIL PHARM INC.
 XX Darrow AL, Qi J, Andrade-Gordon P;
 XX WPI; 2002-106601/14.
 XX P-PSDB; AAG78577.
 XX Nucleic acid encoding a serine protease called D-G protein which is
 FT useful for identifying modulators that are useful for treating a
 PT condition which is mediated by protease D-G, e.g. cancer, skin
 PT disorders, or neuropathic pain -
 XX Claim 4; Fig 1A; 81pp; English.
 XX The invention relates to an isolated and purified nucleic acid that
 CC encodes a serine protease called D-G protein. The activity of the protein
 CC of the invention may be described as cytostatic, anti-inflammatory,
 CC dermatological and anticoagulation. The serine protease of the invention
 CC is a member of the trypsin/chymotrypsin-like (SI) serine protease family,
 CC which play an important role in processes such as digestion and
 CC regulatory amplification cascades through the proteolytic activation of
 CC inactive zymogen precursors. Protease D-G modulating compounds are useful
 CC for treating a condition which is mediated by protease D-G, e.g. cancer,

C skin disorders, neuropathic pain, inflammatory disorders, or coagulation
 C diathesis/thrombosis. The polynucleotide encoding the protease is useful
 C for identifying, detecting or isolating mutant forms of DNA molecules
 C encoding the protease. The protease is useful for identifying modulators
 C of the functional protease. The D-G protein can be used for formulation
 C of compositions for laundry detergents and skin care products. Protease
 C D-G gene therapy may be used to introduce protease D-G into the cells of
 C target organisms. As the D-G protein is derived from a human, it is less
 C likely to produce an allergic reaction in sensitive individuals when used
 C in formulations for laundry detergents and skin care products. The
 C current sequence represents the human serine protease D-G cDNA.

X Sequence 2121 BP; 487 A; 614 C; 584 G; 436 T; 0 other;

Alignment Scores:

red. No.: 5.48e-209 Length: 2121
 core: 2395.00 Matches: 434
 Percent Similarity: 99.77% Conservative: 0
 Best Local Similarity: 99.77% Mismatches: 1
 Query Match: 99.70% Indels: 0
 B: 24 Gaps: 0

S-09-607-745-2 (1-435) x AAI64284 (1-2121)

1 MetAspProAspSerAspGlnProLeuAsnSerLeuAspValLysProLeuArgLysPro 20
 277 ATGATCCTGACAGTGATCAACCTCTGAACAGCTCGATGTCAACCCCTGGCGAAACCC 336
 21 ArgIleProMetGluThrPheArgLysValGlyLeuProIleIleLeuLeuLeuSer 40
 337 CGTATCCCATGAGACCTTCAGAAAGTGGGATCCCATCATCATAGCACACTGAGC 396
 41 LeuAlaSerIleIleValValValValLeuIleLysValIleLeuAspLysTyrPhe 60
 397 CTGGCAGATATCATCATGTGGTGTCTCATCAAGGTGATCTGGATAAATACTTTC 456
 61 LeuCysGlyGlnProLeuHisPheIleProArgLysGlnLeuCysAspGlyGluLeuAsp 80
 457 CTCTGGGGAGCTCTCCACTTCATCCGAGGAGCAGCTGTGTGACGGAGAGCTGGAC 516
 81 CysProLeuGlyGluAspGluHisCysValLysSerPheProGluGlyProAlaVal 100
 517 TGTCTCCCTTGGGGAGGACGAGGACACTGTGTCAAGAGCTTCCCGAAGGCGCTGCAATG 576
 101 AlaValArgLeuSerLysAspArgSerThrIleuGlnValLeuAspSerAlaThrGlyAsn 120
 577 GCAGTCCGCTCTCCAAAGGACCGATCCACATCGAGTGTGGAGCTGGCCACAGGGAAC 636
 121 TrpPheSerAlaCysPheAspAsnThrGluAlaLeuAlaGluThrAlaCysArgGln 140
 637 TGTCTCTGCTGTTCGACAACTTCACAGAGCTCTCCCTGAGACAGCCTGTAGGACAG 696
 141 MetGlyTyrSerSerLysProThrPheArgAlaValGluLeuGlyProAspGlnAspLeu 160
 697 ATGGGCTACAGCAGCAACCCACTTTCAGAGCTGTGGAGATTGGCCACAGACAGGATCTG 756
 161 AspValValGluIleThrGluAsnSerGlnGluLeuArgMetArgAsnSerSerGlyPro 180
 757 GATGTGTGTAATATCAGAAAAACAGCAGAGCTTTCGATGCGGAACCTCAAGTGGGCCC 816
 181 CysLeuSerGlySerLeuValSerLeuHisCysLeuAlaCysGlyLysSerLeuLysThr 200
 817 TGTCTCTCAGGCTCCCTGGTCTCCCTGCATGTCTTCCCTGTGGAGAGAGCCTGAGAGACC 876
 201 ProArgValValGlyGlyGluGluAlaSerValAspSerTrpProTrpGlnValSerIle 220
 877 CCCCGTGTGGTGGGAGGAGGCGCTCTGTGATCTTCTGGCTTGGCAGAGTCAAGATC 936
 221 GlnTyrAspLysGlnHisValCysGlySerIleLeuAspProHisTrpValLeuThr 240
 937 CAGTACGACAAACAGCAGCTCTGTGGAGGAGGATCCCTGGACCCCACTGGGTCTCAGC 996
 241 AlaAlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySer 260

Db 997 GCAGCCCACTGCTTCAGGAAACATACCGATGTGTTCAACTGGAAGTGCAGGAGCTCA 1056
 QY 261 AspLysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleIleGluPheAsn 280
 Db 1057 GACAAACTGGGAGCTTCCATCCCTGGCTGTGGCCAGATCATCATTAATTAATCAAC 1116
 QY 281 ProMetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPhe 300
 Db 1117 CCCATGTACCCCAACACAAATGATGACATGCGCCTCATGAAGTGCAGTTCCTCACTCCTTC 1176
 QY 301 SerGlyThrValArgProIleCysLeuProPhePheAspGluLeuThrProAlaThr 320
 Db 1177 TCAGGCACAGTACAGGCCCATCTGTCTGCCCTTCTTTGATGAGGAGCTCACTCCAGCCACC 1236
 QY 321 ProLeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyLysMetSerAspIle 340
 Db 1237 CCACTCTGGATCATTGGATGGGGCTTTACGAGCAGAAATGGAGGAGATGTCTGACATA 1296
 QY 341 LeuLeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAlaTyr 360
 Db 1297 CTGCTGACAGCGCTCAGTCCAGGTCAATTGACAGCACACGGTGCATTCAGACGATGGGTAC 1356
 QY 361 GlnGlyGluValThrGluLysMetCysAlaGlyIleProGluGlyValAspThr 380
 Db 1357 CTGGGGAGAGTACCCGAGAGATGATGTGTGAGGATCCCGAAGGGGTGTGGACACC 1416
 QY 381 CysGlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGly 400
 Db 1417 TGCCAGGGTGCAGAGTGTGGGCCCCCTGATGTACCAATCTTGACAGTGGCANTGTGGTGGC 1476
 QY 401 IleValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyValTyrThrLysVal 420
 Db 1477 ATCGTTAGTGGGCTATGCTCGGGGGCCCGAGCACCCCGAGGGGTATACACCAAGGTC 1536
 QY 421 SerAlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu 435
 Db 1537 TCAGCCTATCTCAACTGGATCTACAAATGCTGGAGGCTGAGCTG 1581
 RESULT 7
 AAD02557
 ID AAD02557 standard; DNA; 1479 BP.
 XX AAD02557;
 AC AAD02557;
 XX 02-MAY-2001 (first entry)
 DT 02-MAY-2001 (first entry)
 XX Human seripancrin variant #1 DNA.
 DE Human seripancrin variant #1 DNA.
 XX Human; seripancrin; serine protease; chromosome 11q22-q23; therapy;
 KW arthritis; chronic obstructive pulmonary disorder; COPD; cancer;
 KW osteoporosis; aberrant wound healing; angiogenesis; diabetes;
 KW inflammatory disorder; stroke; cardiovascular disease; gene therapy;
 KW vaccine; cytosstatic; cerebroprotective; vulnery; osteopathic; ds.
 XX Homo sapiens.
 OS Homo sapiens.
 XX Key
 FH Location/Qualifiers
 FT 1..1479
 CDS /tag= a
 FT /product= "Human seripancrin variant #1 protein"
 FT
 FT
 FT
 XX WO200104141-A2.
 PN 18-JAN-2001.
 XX 04-JUL-2000; 2000WO-BP06211.
 PF 12-JUL-1999; 99EP-0113428.
 XX (MERE) MERCK PATENT GMBH.
 XX Suendermann B, Hofmann U, Matzku S, Wilbert O;
 PI

WPI; 2001-147177/15.
P-PSDB; AAY72559.

P-PSDB; AAY72559.

New extracellular serine protease Seripancrin, useful for treating cancer, osteoporosis, arthritis, chronic obstructive pulmonary disease, diabetes, inflammatory disorders, stroke, angiogenesis and aberrant wound healing -

Disclosure: Page 40-42; 45pp; English.

The present invention relates to seripancrin polynucleotides, and polypeptides encoded by them. Seripancrin are members of serine protease family. This protein contains a transmembrane domain, a low density lipoprotein (LDL) domain, protease domain and a scavenger receptor cysteine-rich (SRCR) domain. The LDL and SRCR domains help to define the specificity of seripancrin's intra and intermolecular interactions. The polynucleotides and polypeptides of the invention are useful for treating and diagnosing diseases such as arthritis, chronic obstructive pulmonary disorder (COPD), cancer, osteoporosis, aberrant wound healing, angiogenesis, inflammatory disorders, diabetes, stroke and cardiovascular diseases. Seripancrin genes are useful in chromosome localisation studies, as tools for tissue expression studies and also in gene therapy. The polypeptides of the invention are used for identifying agonists and antagonists useful for treating conditions associated with seripancrin imbalance. These polypeptides are also useful as vaccines. The present sequence is a DNA coding for seripancrin variant #1 protein. The seripancrin gene is located on human chromosome 11q22-q22.

Sequence 1479 BP: 329 A; 425 C; 407 G; 318 T; 0 other;

ament scores:

Length:	1.24e-208	1479
Matches:	2329.00	432
Conservative:	100.00%	0
Mismatches:	100.00%	0
Indels:	99.44%	0
Gaps:	22	0

9-607-745-2 (1-435) x AAD02557 (1-1479)

[illegible]

PERCUTANEOUS

RESULTS
ABZ35523

AB235523
ID AB235523 standard: cDNA: 2165 BP.

XX
AT

AC ABZ35523;

XX

DT 05-FEB-2003 (first entry)

XX

DE Human gene expression profile polynucleotide SEQ ID NO 634.

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

KW Human; artery; endothelium

KW bronchial epi

50

WO200274979-A2.
 26-SEP-2002.
 20-MAR-2002; 2002WO-US08456.
 20-MAR-2001; 2001US-276947P.
 (ORTH) ORTHO CLINICAL DIAGNOSTICS INC.
 Wan J, Wang Y;
 WPI; 2002-740862/80.
 New gene expression profile generated from primary, endothelial, epithelial, and muscle cell types, useful for identifying disease pathologies involving alterations of gene expression, e.g. cancer -
 Disclosure; Page 759-760; 850pp; English.
 The invention relates to a gene expression profile comprising one or more genes (ABZ34889-ABZ35692) and generated from a cell type. The cell type is a coronary artery endothelium, umbilical artery or vein endothelium, aortic endothelium, dermal microvascular endothelium, pulmonary artery endothelium, myometrium microvascular endothelium, keratinocyte epithelium, bronchial epithelium, mammary epithelium, prostate epithelium, renal cortical epithelium, renal proximal tubule epithelium, small airway epithelium, renal epithelium, umbilical artery smooth muscle, neonatal dermal fibroblast, pulmonary artery smooth muscle, dermal fibroblast, neural progenitor cells, skeletal muscle, astrocytes, aortic smooth muscle, mesangial cells, coronary artery smooth muscle, bronchial smooth muscle, uterine smooth muscle, lung fibroblast, osteoblasts or prostate stromal cell. The gene expression profile is used for determining the level of RNA expression for a sample, determining the phenotype of a cell and distinguishing cell types. The gene or a protein expression profile is useful in identifying disease pathologies involving alterations of gene expression. The assessment of expression profiles may provide meaningful information with respect to tumour type and stage, treatment methods, and prognosis. The gene or protein expression profile may also be used for creating microarrays. The microarray is useful for genetic and physical mapping of genomes, DNA sequencing, genetic or medical diagnosis, genotyping of organisms, confirming cell or tissue identifications and in identifying promising antibiotics, antiviral or antifungal agents.
 Sequence 2165 BP; 502 A; 623 C; 595 G; 445 T; 0 other;
 Alignment Scores:
 d. No.: 2,06e-208 Length: 2165
 re: 2329.00 Matches: 435
 cent Similarity: 99.77% Conservative: 0
 t Local Similarity: 99.77% Mismatches: 0
 ry Match: 99.44% Indels: 1
 24 Gaps: 0
 09-607-745-2 (1-435) x ABZ35523 (1-2165)
 1 MetAspProAspSerAspGlnProLeuAnSerLeuAspValLysProLeuArgLysPro 20
 310 ATGGATCCCTGACAGTATCAACCTCTGAACAGCCTCGATGTCAAAACCCCTGGCAACCC 369
 21 ArgileProMetGluThrPheArgLysValGlyIleProIleIleIleAlaLeuLeuSer 40
 370 CGTATCCCATGAGACCTTCAGAAAGGTGGGATCCCATATCATAGACCTACTGAGC 429
 41 LeuAlaSerIleIleIleValValValLeuIleLysValIleLeuAspLysTyrIlePhe 60
 430 CTGGCAGTATCATCTTGTGTGTCTCATCAAGGTGATCTTGATTAATACTACTTTC 489
 61 LeuCysGlyGlnProLeuHisPheIleProArgLysGlnLeuCysAspGlyGluLeuAsp 80
 490 CTCTCGGGCAGCCTCTCCACTTCATCCGAGGAGCAGCTGTGTACGAGGAGCTGGAC 549

1493 AGCCTATCTCACTGCACTACATGCTCGAGAGCTGAGCTG 1525

db

RESULT 10
AADI3169
AADI3169 standard; DNA; 2079 BP.
AADI3169;
16-OCT-2001 (first entry)
Human transmembrane protease serine 4 (TMPRSS4) DNA.
Human; transmembrane serine protease; membrane-type serine protease;
MTSP; protease domain; neoplastic disease; tumour; cancer; cytostatic;
lung carcinoma; colon adenocarcinoma; ovarian carcinoma; gene therapy;
TMPRSS4; ds.
Homo sapiens.

Key Location/Qualifiers
CDS 251..1522
/tag= a
/product= "Human transmembrane protease, serine 4
(TMPRSS4)"

WO200157194-A2.
09-AUG-2001.
02-FEB-2001; 2001WO-US03471.
03-FEB-2000; 2000US-0179982.
18-FEB-2000; 2000US-0183542.
22-JUN-2000; 2000US-0213124.
26-JUL-2000; 2000US-0220970.
08-SEP-2000; 2000US-0657986.
22-SEP-2000; 2000US-0234840.
(CORV-) CORVAS INT INC.

Madison EL, Ong EO, Yeh J;
WPI; 2001-488877/53.
P-PSDB; AAE06944.
Novel single chain polypeptide comprising protease domain of type-II
membrane-type serine protease or its catalytically active portion
useful for treating and preventing cancer and tumor
Disclosure; Page 253-255; 256pp; English.

The invention relates to transmembrane serine proteases and their
corresponding nucleotides and the protease domain of a type-II
membrane-type serine protease (MTSP). MTSP is useful for identifying
compounds that modulate or inhibit its proteolytic activity and for
formulating a medicament for treating neoplastic disease. MTSP and
its corresponding nucleotides are useful in preventing or treating
tumors or cancers such as lung carcinoma, colon adenocarcinoma and
ovarian carcinoma, in diagnostics and in hybridisation assays. MTSP
is useful as a diagnostic marker for tumour development, growth and/or
progression and as immunogens to generate antibodies that specifically
bind to it. MTSP DNA is useful in a yeast two-hybrid system and in
gene therapy. The present sequence is a DNA encoding human
transmembrane protease, serine 4 (TMPRSS4) protein.

Sequence 2079 BP; 489 A; 594 C; 575 G; 421 T; 0 other;

Alignment Scores:

Seq. No.: 5.74e-208 Length: 2079
core: 2324.00 Matches: 434
Percent Similarity: 99.77% Conservative: 0
Best Local Similarity: 99.77% Mismatches: 0
Query Match: 99.23% Indels: 1

DB:	22	Gaps:	0
US-09-607-745-2 (1-435) x AADI3169 (1-2079)			
QY	2	AspProAspSerAspGlnProLeuAsnSerLeuAspValLysProLeuArgLysProArg	21
Db	217	GATCCTGACAGTGATCAACCTCTGAACAGCTCGATGTCAAACCCCTGCGCAACCCCGT	276
QY	22	IleProMetGluThrPheArgLys-ValGlyIleProIleIleIleAlaLeuLeuSerLe	41
Db	277	ATCCCATGAGACCTTCAGAAAGTGTGGGATCCCATCATCATAGCACTACTGACCT	336
QY	41	uAlaSerIleIleIleValValValLeuIleLysValIleLeuAspLysTyrTyrPheLe	61
Db	337	GGCAGATATCATATTGTGTCTCATCAAGGTGATTCTGGATAAATACTACTTCTCT	396
QY	61	uCysGlyGlnProLeuHisPheIleProArgLysGlnLeuCysAspGlyGlnLeuAspCy	81
Db	397	CTGGGGCAGCCTCTCCACTTCATCCGAGAGACGACTGTGTGACGAGAGCTGGACTG	456
QY	81	sProLeuGlyGluAspGluGluHisCysValLysSerPheProGluGlyProAlaValAl	101
Db	457	TCCCTTGGGGAGGACGAGAGACACTGTGTCAAGAGCTTCCCGAAGGGCTGCACTGGC	516
QY	101	aValArgLeuSerLysAspArgSerThrLeuGlnValLeuAspSerAlaThrGlyAsnTr	121
Db	517	AGTCCGCTCTCCAGGACCGATCCACTGTCAGGTGCTGGACTCGGCCACAGGAACTG	576
QY	121	pPheSerAlaCysPheAspAsnPheThrGluAlaLeuAlaGluThrAlaCysArgGlnMe	141
Db	577	GTTCCTCTGCCCTGTTTCGACAACTTCAGAAAGCTCTCGCTGACAGAGCTGTAGGCAGAT	636
QY	141	tGlyTyrSerSerLysProThrPheArgAlaValGluIleGlyProAspGlnAspLeuAs	161
Db	637	GGCTACAGCAGCAAAACCCACTTTCAGAGCTGTGGAGATGGCCACAGACAGATCTGGA	696
QY	161	pValValGluIleThrGluAsnSerGlnGluLeuArgMetArgAsnSerSerGlyProCy	181
Db	697	TGTTGTTGAAATCACAGAAACACAGCAGGAGCTTCGCATGCGGAACTCAAGTGGGCCCTG	756
QY	181	sLeuSerGlySerLeuValSerLeuHisCysLeuAlaCysGlyLysSerLeuLysThrPr	201
Db	757	TCCTCAGGCTCCCTGCTCTCCCTGCACTGTCTGTGCTGTGGGAGAGCCTGAAGACCC	816
QY	201	oArgValValGlyGlyGluAlaSerValAspSerTrpProTyrGlnValSerIleGl	221
Db	817	CCGTGTGTGGTGGGAGAGAGCCCTCTGTGATTTCTTGGCCTTGGCAGGTGAGATCCA	876
QY	221	nTyrAspLysGlnHisValCysGlySerIleLeuAspProHisTrpValLeuThrAl	241
Db	877	GTACGACAAACAGCAGCTCTGTGGAGGAGCATCTCGACCCCTCTGGTCTCTACGGC	936
QY	241	alaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySerAs	261
Db	937	AGCCCACTCTTTCAGAAACATACCGATGTTCACCTGGAAGGTGCGGCGAGCTCAGA	996
QY	261	pLysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleGluPheAsnPr	281
Db	997	CAAACTGGGAGCTTCCCATCCCTGGCTGTGGCCAGATCATCATTTGAATTCACCC	1056
QY	281	oMetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPheSe	301
Db	1057	CATGTACCCCAAGACATGACATGCCCTCATGAAGCTGCAGTTCCTCCACTCATTCTC	1116
QY	301	rgLysThrValArgProIleCysLeuProPhePheAspGluGluLeuThrProAlaThrPr	321
Db	1117	AGGCACAGTCRGGCCCATCTGTCTGCCCTTCTTTCATGAGGAGCTCACTCCAGCCACC	1176
QY	321	oLeuTyrIleIleGlyTyrGlyPheThrLysGlnAsnGlyGlyLysMetSerAspIleLe	341
Db	1177	ACTCTGGATCATTTGATGGGCTTTTACGAAGCAGATGGAGGAGAGATGCTGACATCT	1236
QY	341	uLeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAspAlaTyrGl	361

```

|||||
1237 GCTCAGCGCTCAGTCCAGGTCATCTGACAGACACACGCGTCAATGACGATCGGTACCA 1296
|||||
361 nGlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyGlyValAspThrCy 381
|||||
1297 GGGGGAAGTACCCGAGAAGATGATGTGACGAGCATCCCGAAGGGGTGTGGACACCTG 1356
|||||
381 sGlnGlyAspSerGlyGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGlyI1 401
|||||
1357 CCAGGGTGACAGTGGTGGGCCCTGATGTACCAATCTGACCAATCTGACCAATCTGATG 1416
|||||
401 eValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyValTyrThrIysValSe 421
|||||
1417 CGTTAGCTGGGGCTATGGCTCGGGGGCCCGAGACCCCGAGAGTATACACCAAGGTCTC 1476
|||||
421 rAlaTyrLeuAsnTrpIleTyrAsnValTyrLysAlaGluLeu 435
|||||
1477 AGCTATCTCACTGATCTACATATGCTGGAGGCTGAGCTG 1519
|||||
RESULT 11
AI72975
D AAI72975 standard; cDNA; 2079 BP.
X
X AAI72975;
X
X 21-AUG-2002 (first entry)
X CUA8 cDNA.
X
X Gene; colorectal cancer; CGA7; CUA8; modulating protein; screening;
X drug candidate; vaccine; ss.
X
X Homo sapiens.
X
X Key Location/Qualifiers
X CDS /*tag= a
X /product= "CUA8"
X
X US2002042067-A1.
X
X 11-APR-2002.
X
X 08-MAY-2001; 2001US-0851588.
X
X 17-AUG-2000; 2000US-0642252.
X 06-SEP-2000; 2000US-0656002.
X
X (MACK/) MACK D.
X (GISH/) GISH K C.
X (WILS/) WILSON K E.
X
X Mack D, Gish KC, Wilson KE;
X
X WPI; 2002-453647/48.
X P-PSDB; AAG79358.
X
X Screening drug candidates for treating colorectal cancer, comprises
X determining the effect of the candidate on the expression profile gene
X of CGA7 or CUA8 -
X
X Disclosure; Fig 5; 40pp; English.

```

CC candidate colorectal cancer drug. An antibody to CGA7 or CUA8 can inhibit the activity of CGA7 or CUA8, respectively, and is used to screen for an agent that can interfere with the binding of CGA7 or CUA8 to the antibody. The antibody can be used to treat colorectal cancer. The antibody or a fragment of it is used to localize a therapeutic group to a colorectal cancer tissue, where the therapeutic group is a cytotoxic agent or a radioisotope. Antisense molecules are used to inhibit colorectal cancer in a cell. Nucleic acid segments encoding CGA7 or CUA8 are used in a biochip. CGA7, CUA8 or a nucleic acid encoding it are used to elicit an immune response. CGA7 or CUA8 is used to determine the prognosis of an individual with colorectal cancer. CC Nucleic acid encoding CGA7 or CUA8 can be used in vaccines.

XX Sequence 2079 BP; 489 A; 594 C; 575 G; 421 T; 0 other;

Alignment Scores:

Pred. No.:	5,74e-208	Length:	2079
Score:	2324.00	Matches:	434
Percent Similarity:	99.77%	Conservative:	0
Best Local Similarity:	99.77%	Mismatches:	1
Query Match:	99.23%	Indels:	1
DB:	24	Gaps:	0

US-09-607-745-2 (1-435) x AAI72975 (1-2079)

QY	2 AspProAspSerAspGlnProLeuAsnSerLeuAspValIysProLeuArgLysProArg	21
DB	217 GATCTCAGCAGTATCAACCTCTGACAGGCTCGATGTCAAAACCCCTGCGCAACCCCGT	276
QY	22 IleProMetGluThrPheArgLys-ValGlyIleProIleIleIleAlaLeuLeuSerLe	41
DB	277 ATCCCATGAGACCTTCAGAAAGTGGGGATGCCATCATCATCATCATCATCATCATCAT	336
QY	41 uAlaSerIleIleIleValValValValIleLysValIleLeuAspLysTyrTrpPheLe	61
DB	337 GGCAGTATCATCATTTGGTTGCTTCATCAGGTGATTCGATTAATACTACTTCTCT	396
QY	61 uCysGlyGlnProLeuHisPheIleProArgLysGlnLeuCysAspGlyGluLeuAspCy	81
DB	397 CTGGGGGAGCCTCTCCACTTCATCCGAGGAGCAGCTGTGTGACGAGAGTGGACTG	456
QY	81 sProLeuGlyGluAspGluGluHisCysValLysSerPheProGluGlyProAlaValAl	101
DB	457 TCCCTTGGGGGAGGAGGAGGAGCCTGTGTCAAGAGCTTCCCGAAGGGCTGCAGTGGC	516
QY	101 aValArgLeuSerLysAspArgSerThrLeuGlnValLeuAspSerAlaThrGlyAsnTr	121
DB	517 AGTCGGCTCTCCAAAGGACCGATCCACACTGCGAGGTCTGACTCGGCCACAGGGAATG	576
QY	121 pPheSerAlaCysPheAspAsnPheThrGluAlaLeuAlaGluThrAlaCysArgGlnMe	141
DB	577 GTTCTCTGCTGTTTCGACAACTTCACAGAGCTCTCGCTGAGACACCTGTAGCGAGAT	636
QY	141 tGlyTyrSerSerLysProThrPheArgAlaValGluIleGlyProAspGlnAspLeuAs	161
DB	637 GGGCTACAGCAGCAAAACCCACTTTTCAGAGCTGTGGAGATTGGCCACAGAGATCTGGA	696
QY	161 pValValGluIleThrGluAsnSerGlnGluLeuArgMetArgAsnSerSerGlyProCy	181
DB	697 TGTGTGTGAATCATCAGAAACACCGAGGAGCTTCGCATCGGGAACCTCAAGTGGGGCTG	756
QY	181 sLeuSerGlySerLeuValSerLeuHisCysLeuAlaCysGlyLysSerLeuLysThrPr	201
DB	757 TCTCTCAGGCTCTCTGGTCTCCCTGCACTGTCTGTCTGTGGGAGAGGCTTGAGACCCC	816
QY	201 oArgValValGlyGlyGluGluAlaSerValAspSerTrpProTrpGlnValSerIleGl	221
DB	817 CCGTGTGTGGGTGGGAGGAGGCTCTGTGTGATTTCTTGGCTTGGCAGGTGACATCCA	876
QY	221 nTyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThrAl	241
DB	877 GTACACAAACAGCAGCTCTGTGTGGAGGAGCATCTCTGACCCCATCTGGGTCTCTCAGGC	936

241 aAlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySerAs 261
 937 AGCCACCTGCTTCCAGGAACATACCGATGCTTCAACTGGAAGTGGCGGAGCTCAGA 996
 261 pLysLeuGlySerPheProSerLeuAlaValAlaLysLeilelleleGluPheAsnPr 281
 997 CAACAGTGGGAGCTTCCATCCCTGGCTGGTGGCAGATCATCATGTAATTCACCC 1056
 281 oMetTyProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPheSe 301
 1057 CATGTACCCCAAGACAAATGACATCGCCCTCATGAAGCTGCAGTCCCACTCTTCTC 1116
 301 rGlyThrValArgProIleCysLeuProPhePheAspGluGluLeuThrProAlaThrPr 321
 1117 AGGCACAGTACAGGCCCATCTGCTGCGCTTCTTTGATGAGAGAGCTCACTCCAGCACCCC 1176
 321 oLeuTrpIlelleGlyTrpGlyPheThrLysGlnAsnGlyLysMetSerAspIleLe 341
 1177 ACTCTGATCATTTGGATGGGCTTTAGAGCAGCAATGGAGGAGATGTCTGACACT 1236
 341 uLeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAlaTrpGI 361
 1237 GCTGCAGCGCTCAGTCCAGGTCATTGACAGCACACGCTGCAATGCACAGCATGCTACCA 1296
 361 nGlyGluValThrGluLysMetMetCysAlaGlyIleProGluGlyValAspThrCy 381
 1297 GGGGGAAGTACCAGAGAGATGATGTGTGAGGAGCATCCCGAAGGGGTGTGGACACCTG 1356
 381 sGlnGlyAspSerGlyProLeuMetTyTrpGlnSerAspGlnTrpHisValValGlyI 401
 1357 CCAGGTGACAGTGTGGGCCCTTGATGATGATGATGATGATGATGATGATGATGATGAT 1416
 401 eValSerTrpGlyTyGlyCysGlyProSerThrProGlyValThrLysValSe 421
 1417 CGTAGCTGGGCTATGGCTGCGGGGCGGAGCAGCCAGGAGTATACCAAGGTCTC 1476
 421 rAlaTyLeuAsnTrpIleTyTrpAsnValTrpLysAlaGluLeu 435
 1477 AGCCTATCTCACTGGATCTACATGTCTGAGAGGCTGAGCTG 1519
 ESUL 12
 BS57763
 D ABSS57763 standard; cDNA; 2079 BP.
 C ABSS57763;
 X 05-FEB-2003 (first entry)
 X cDNA encoding human colorectal cancer modulating protein CUA8.
 X Colorectal cancer modulating protein; BCMP; CUA8; colorectal cancer;
 W gene therapy; human; gene; ss.
 S Homo sapiens.
 S
 S Location/Qualifiers
 H 251..1522
 T /*tag= a
 T /product= "CUA8"
 T /note= "Colorectal cancer modulating protein"
 T /trans_except= (pos:314..316, aa:Leu)
 X
 X US6455668-B1.
 X
 X 24-SEP-2002.
 X
 X 06-SEP-2000; 2000US-0656002.
 X
 X 15-MAR-2000; 2000US-0525993.
 X
 X 28-JAN-2000; 2000US-0493444.
 X
 X (BOS-) EOS BIOTECHNOLOGY INC.

PI Mack D, Gish KC, Wilson KE;
 XX WPI; 2003-066245/06.
 DR P-PSDB; ABG72428.
 XX
 PT Diagnosing colorectal cancer comprises determining the expression of a
 PT gene encoding CUA8 in a first colon tissue of a first individual, and
 PT comparing them to the expression of a gene encoding CUA8 in a second
 PT normal tissue
 XX
 PS Claim 11; Fig 1; 3lpp; English.
 XX
 CC The invention describes a method of diagnosing colorectal cancer
 CC comprising determining the expression of a gene encoding CUA8 or its
 CC fragment in a first colon tissue of an individual, and comparing the
 CC expression of the gene to the expression of the gene in normal tissue
 CC from the individual or in tissue from a second unaffected individual.
 CC A difference in the expression indicates that the first individual has
 CC colorectal cancer. The CUA8 genes and proteins are useful for diagnosing
 CC and prognosticating colorectal cancer. The methods are also useful for
 CC screening candidate bioactive agents that can modulate colorectal
 CC cancer, or for treating or inhibiting colorectal cancer in a patient
 CC e.g. by gene therapy. This sequence encodes the human colorectal
 CC cancer modulating protein CUA8.
 XX
 SQ Sequence 2079 BP; 489 A; 594 C; 575 G; 421 T; 0 other;

Alignment Scores:

Pred. No.: 5,74e-208 Length: 2079
 Score: 2224.00 Matches: 434
 Percent Similarity: 99.77% Conservative: 0
 Best Local Similarity: 99.77% Mismatches: 0
 Query Match: 99.23% Indels: 1
 DB: 25 Gaps: 0

US-09-607-745-2 (1-435) x ABSS57763 (1-2079)

QY 2 AspProAspSerAspGlnProLeuAsnSerLeuAspValLysProLeuArgLysProArg 21
 Db 217 GATCTTGACAGTATCACTCTGACAGCTCGATGTCAAACCCCTGCGCAACCCCGT 276
 QY 22 IleProMetGluThrPheArgLys-ValGlyIleProIleIleAlaLeuLeuSerLe 41
 Db 277 ATCCCATGGAGACCTTCAGAAAGTGTGGGATGCCCATCATCATAGCACTACTAGCCT 336
 QY 41 uAlaSerIleIleIleValValValLeuIleLysValIleLeuAspLysTyTrpPheLe 61
 Db 337 GGCGAGTATCACTGTGTGTCTCATCAAGGTGATTCGGATAAATACTACTTCTCT 336
 QY 61 uCysGlyGlnProLeuHisPheIleProArgLysGlnLeuCysAspGlyLysLeuAspCy 81
 Db 397 CTGGGGCAGCCTCTCCACTTCATCCGAGGAGCAGCTGTGTGACGAGAGCTGGACTG 456
 QY 81 sProLeuGlyGluAspGluHisCysValLysSerPheProGluGlyProAlaValAl 101
 Db 457 TCCCTTGGGGAGGACGAGGAGCACTGTCTCAAGAGCTTCCCGAAGGCGCTGACTGGC 516
 QY 101 aValArgLeuSerLysAspArgSerThrLeuGlnValLeuAspSerAlaThrGlyAsnTr 121
 Db 517 AGTCCGCTCTTCAGGACCGATCCACACTGAGGTGTGGACTCGGCCACAGGGAAGT 576
 QY 121 pPheSerAlaCysPheAspAsnPheThrGluAlaLeuAlaGluThrAlaCysArgGlnMe 141
 Db 577 GTTCTCTGCTGTTCGACAACTTCACAGAACTCTGCTGAGACAGCTCTTAGGAGAT 636
 QY 141 tGlyTySerSerLysProThrPheArgAlaValGluIleGlyProAspGlnAspLeuAs 161
 Db 637 GGGCTACAGCAGCAAAACCCACTTTCAGAGCTGTGGAGATTGGCCAGACAGGATCTGGA 696
 QY 161 pValValGluIleThrGluAsnSerClnGluLeuArgMetArgAsnSerSerGlyProCy 181
 Db 697 TGTGTGTGAATTCACAGAAACAGCAGGAGCTTCGATGCGGAACTCAAGTGGGCCCTG 756

30-SEP-1998;	98US-0102494;
30-SEP-1998;	98US-0102487;
30-SEP-1998;	98US-0102570;
30-SEP-1998;	98US-0102571;
01-OCT-1998;	98US-0102664;
01-OCT-1998;	98US-0102687;
02-OCT-1998;	98US-0102965;
06-OCT-1998;	98US-0103258;
06-OCT-1998;	98US-0103449;
07-OCT-1998;	98US-0103314;
07-OCT-1998;	98US-0103315;
07-OCT-1998;	98US-0103328;
07-OCT-1998;	98US-0103395;
07-OCT-1998;	98US-0103396;
07-OCT-1998;	98US-0103401;
08-OCT-1998;	98US-0103633;
08-OCT-1998;	98US-0103678;
08-OCT-1998;	98US-0103679;
08-OCT-1998;	98US-0103711;
14-OCT-1998;	98US-0104257;
20-OCT-1998;	98US-0104987;
20-OCT-1998;	98US-0105000;
20-OCT-1998;	98US-0105002;
21-OCT-1998;	98US-0105104;
22-OCT-1998;	98US-0105169;
22-OCT-1998;	98US-0105266;
26-OCT-1998;	98US-0105693;
26-OCT-1998;	98US-0105694;
27-OCT-1998;	98US-0105807;
27-OCT-1998;	98US-0105881;
27-OCT-1998;	98US-0105882;
27-OCT-1998;	98US-0106062;
28-OCT-1998;	98US-0106023;
28-OCT-1998;	98US-0106030;
28-OCT-1998;	98US-0106032;
28-OCT-1998;	98US-0106033;
28-OCT-1998;	98US-0106178;
29-OCT-1998;	98US-0106248;
29-OCT-1998;	98US-0106384;
29-OCT-1998;	98US-0108500;
03-NOV-1998;	98US-0106856;
03-NOV-1998;	98US-0106902;
03-NOV-1998;	98US-0106905;
03-NOV-1998;	98US-0106919;
03-NOV-1998;	98US-0106932;
03-NOV-1998;	98US-0106934;
10-NOV-1998;	98US-0107783;
17-NOV-1998;	98US-0108779;
17-NOV-1998;	98US-0108787;
17-NOV-1998;	98US-0108788;
17-NOV-1998;	98US-0108801;
17-NOV-1998;	98US-0108802;
17-NOV-1998;	98US-0108806;
17-NOV-1998;	98US-0108807;
17-NOV-1998;	98US-0108867;
17-NOV-1998;	98US-0108925;
18-NOV-1998;	98US-0108848;
18-NOV-1998;	98US-0108849;
18-NOV-1998;	98US-0108850;
18-NOV-1998;	98US-0108851;
18-NOV-1998;	98US-0108858;
18-NOV-1998;	98US-0108904;

XX	New mammalian DNA sequences encoding transmembrane, receptor or
PT	secreted PRO polypeptides, useful for screening of potential peptide or
PT	small molecule inhibitors of the relevant receptor/ligand interactions
XX	
XX	Claim 2; Fig 155; 773pp; English.
XX	
AA	AA37022 to AA37144 encode the new isolated human transmembrane,
CC	receptor or secreted PRO polypeptides given in AA599340 to AA599462. The
CC	transmembrane and receptor PRO proteins can be used for screening of
CC	potential peptide or small molecule inhibitors of the relevant
CC	receptor/ligand interactions. The polypeptides and nucleotide sequences
CC	encoding them have various industrial applications, including uses as
CC	pharmaceutical and diagnostic agents. AA37145 to AA37330 represent
CC	PCR primers and hybridisation probes used in the isolation of the PRO
CC	polypeptides from the present invention.
XX	
SQ	Sequence 2063 BP; 477 A; 591 C; 576 G; 419 T; 0 other;

Best Local Similarity: 98.85%

```

Every Match: 98.10% Indels: 5
              22      Gaps: 1
3-09-607-745-2 (1-435) x AAS46089 (1-2063)
2 AspProAspSerAspGlnProLeuAsnSerLeuAspValLysProLeuArgLysProArg 21
219 GATCTGTGACAGTGAACAACCTCTGACACGCTCGATGTCAAACTCTGGCGAAACCCCT 278
22 IleProMetGluThrPheArgLysValGlyIleProIleIleIleAlaLeuLeuSerLeu 41
279 ATCCCATCGAGACCTTCAGAAAGGTGGGATCCCATCATCATATGAGCACTACTGAGCTG 338
42 AlaSerIleIleIleValValValLeuIleLysValIleLeuAspLysTyrThrPheLeu 61
339 GCGAGTATCATCATTTGTGTTCTCATCAAGGTGAATCTCGATAAATACTACTTCCTC 398
62 CysGlyGlnProLeuHisPheIleProAlaGlyGlnLeuCysAspGlyGluLeuAspCys 81
399 TCGGGGAGCTCTCCACTTCTATCCGAGAGAGCAGCTGTGTGACGGAGAGCTGGACTGT 458
82 ProLeuGlyGluAspGluHisCysValLysSerPheProGluGlyProAlaValAla 101
459 CCTTGGGGAGGAGCAGAGCAGCTGTGTCAAGAGCTTCCCCGAAAGGGCTCGCAGTGGCA 518
102 ValArgLeuSerLysAspArgSerThrLeuGlnValLeuAspSerAlaThrGlyAsnTrp 121
519 GTCCGCCCTCTCCAGGACCGATCCACACTGCGAGGTGCTGACTCGGCCACAGGGAACTGG 578
122 PheSerAlaCysPheAspAsnPheThrGluAlaLeuAlaGluThrAlaCysArgGlnMet 141
579 TTCTCTGCTGTTTCGACAACTTCACAGAGCTCTGCTGACAGACCTGTAGGAGCAGTG 638
142 GlyTyrSerSerLysProThrPheArgAlaValGluIleGlyProAspGlnAspLeuAsp 161
639 GGCTACAGC-----AGAGCTGTGGAGATTGGGCCACAGCAGGATCTGGAT 683
162 ValValGluIleThrGluAsnSerGlnGluLeuArgMetArgAsnSerSerGlyProCys 181
684 GTTGTGTAAATCACAGAAACACAGGAGCTTCGCATCGGAACCTCAAGTGGGCCCTGT 743
182 LeuSerGlySerLeuValSerLeuHisCysLeuAlaCysGlyLysSerLeuLysThrPro 201
744 CTCTCAGGCTCCCTGGTCTCCCTGCACTCTCTTGTCTGGGAAGAGCTGAGACCCCC 803
202 ArgValValGlyGlyGluGluAlaSerValAspSerTrpProTrpGlnValSerIleGln 221
804 CGTGTGGTGGGTGGGAGGAGGCTCTGTGGATTCTTGGCCTTGGCAGGTCCAGCATCCAG 863
222 TyrAspLysGlnHisValCysGlyGlySerIleLeuAspProHisTrpValLeuThrAla 241
864 TACGACAAACAGCAGCTCTGTGGAGGAGCATCTGGACCCCTGAGTCTCCACGGCA 923
242 AlaHisCysPheArgLysHisThrAspValPheAsnTrpLysValArgAlaGlySerAsp 261
924 GCCCACTGCTTCAGGAACATACCGATGTGTCACTGGAAGGTGGGGCAGGCTCAGAC 983
262 LysLeuGlySerPheProSerLeuAlaValAlaLysIleIleIleGluPheAsnPro 281
984 AAACCTGGGCAGCTTCCCATCCCTGGCTGTGGCCAGATCATCATCATTAATTCAACCCC 1043
282 MetTyrProLysAspAsnAspIleAlaLeuMetLysLeuGlnPheProLeuThrPheSer 301
1044 ATGTACCCCAAGACAATACATCGCCCTCATGAAGCTTCAGTTCCCATCTCTCTCTCA 1103
302 GlyThrValArgProIleCysLeuProPhePheAspGluGluLeuThrProAlaThrPro 321
1104 GGCACAGTCAGGCCCATCTGTCTGCCCTTCTTTGATGAGAGCTCACTCCGCCACCCCA 1163
322 LeuTrpIleIleGlyTrpGlyPheThrLysGlnAsnGlyGlyLysMetSerAspIleLeu 341
1164 CTCTGGATCATTTGGATGGGCTTTACGAGCAGAAATGGAGGAGAGATGTCTGACATCTG 1223

```

Search completed: November 23, 2003, 08:06:43
 Job time : 440.919 secs

```

QY 342 LeuGlnAlaSerValGlnValIleAspSerThrArgCysAsnAlaAspAlaTyrGln 361
DB 1224 CTGCAGGGCTCAGTCCAGTCAATGCAGACACACGGTCAATGCAGACGATCGTACCAG 1283
QY 362 GlyGluValThrGluLysMetCysAlaGlyIleProGluGlyGlyValAspThrCys 381
DB 1284 GGGAGAGTCACCGAGAGATGATGTGTGAGGCATCCCGGAGGGGGTGTGGACACCTGC 1343
QY 382 GlnGlyAspSerGlyProLeuMetTyrGlnSerAspGlnTrpHisValValGlyIle 401
DB 1344 CAGGTGACAGTGTGTGGGCCCTGTATGTATACCAATCTGACCATGTGTGTGGGCATC 1403
QY 402 ValSerTrpGlyTyrGlyCysGlyGlyProSerThrProGlyValTyrThrLysValSer 421
DB 1404 GTTAGCTGGGCTATGGCTGGGGGGCCCGAGCACCCCGAGGATATACACCAAGGTCTCA 1463
QY 422 AlaTyrLeuAsnTrpIleTyrAsnValTrpLysAlaGluLeu 435
DB 1464 GCCTATCTCAACTGGANCTACAATGTCTGGAGGGCTGAGCTG 1505

```